# DETEKSI BERITA PALSU BERBAHASA INDONESIA PADA DATA TIDAK SEIMBANG MENGGUNAKAN *INDOBERT* DENGAN *FOCAL LOSS*

**SKRIPSI**


Oleh :

**AKBAR BIMANTARA T**

**NIM. 220605110080**



**PROGRAM STUDI TEKNIK INFORMATIKA**
**FAKULTAS SAINS DAN TEKNOLOGI**
**UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM**
**MALANG**


**2026**

# DETEKSI BERITA PALSU BERBAHASA INDONESIA PADA DATA TIDAK SEIMBANG MENGGUNAKAN *INDOBERT* DENGAN *FOCAL LOSS*

**SKRIPSI**

Diajukan kepada:

Universitas Islam Negeri Maulana Malik Ibrahim Malang

Untuk memenuhi Salah Satu Persyaratan dalam

Memperoleh Gelar Sarjana Komputer (S.Kom)

**Oleh :**

**AKBAR BIMANTARA T**

**NIM. 220605110080**

**PROGRAM STUDI TEKNIK INFORMATIKA**
**FAKULTAS SAINS DAN TEKNOLOGI**
**UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM MALANG**
**2026**

## HALAMAN PERSETUJUAN

**DETEKSI BERITA PALSU BERBAHASA INDONESIA PADA DATA TIDAK SEIMBANG MENGGUNAKAN *INDOBERT* DENGAN *FOCAL LOSS***

**SKRIPSI**

**Oleh :**
**AKBAR BIMANTARA T**
**NIM. 220605110080**

Telah Diperiksa dan Disetujui untuk Diuji:
Tanggal: 5 Juni 2026

| Pembimbing I, | Pembimbing II, |
|---|---|
| Dr. Zainal Abidin M.Kom, | Nur Fitriyah Ayu Tunjung Sari, M.Cs |
| NIP. 19760613 200501 1 004 | NIP. 19911226 202012 2 001 |

Mengetahui,
Ketua Program Studi Teknik Informatika
Fakultas Sains dan Teknologi
Universitas Islam Negeri Maulana Malik Ibrahim Malang

Supriyono, M. Kom
NIP. 19841010 201903 1 012



iii

## HALAMAN PENGESAHAN

**DETEKSI BERITA PALSU BERBAHASA INDONESIA PADA DATA TIDAK SEIMBANG MENGGUNAKAN *INDOBERT* DENGAN *FOCAL LOSS***

**SKRIPSI**

**Oleh :**
**AKBAR BIMANTARA T**
**NIM. 220605110080**

Telah Dipertahankan di Depan Dewan Penguji Skripsi
dan Dinyatakan Diterima Sebagai Salah Satu Persyaratan
Untuk Memperoleh Gelar Sarjana Komputer ( S.Kom )
Tanggal: 23 Juni 2026

**Susunan Dewan Penguji**

| | | |
|---|---|---|
| Ketua Penguji | : Dr. Ririen Kusumawati, S.Si., M.Kom.<br>NIP. 19720309 200501 2 002 | ( ) |
| Anggota Penguji I | : Fatchurrohman, M.Kom<br>NIP. 19700731 200501 1 002 | ( ) |
| Anggota Penguji II | : Dr. Zainal Abidin, M.Kom<br>NIP. 19760613 200501 1 004 | ( ) |
| Anggota Penguji III | : Nur Fitriyah Ayu Tunjung Sari, M.Cs<br>NIP. 19911226 202012 2 001 | ( ) |

Mengetahui dan Mengesahkan,
Ketua Program Studi Teknik Informatika
Fakultas Sains dan Teknologi
Universitas Islam Negeri Maulana Malik Ibrahim Malang

Supriyono, M.Kom
NIP. 19841010 201903 1 012



iv

## PERNYATAAN KEASLIAN TULISAN

Saya yang bertanda tangan di bawah ini:

| | |
|---|---|
| Nama | : AKBAR BIMANTARA TRAHESTIAWAN |
| NIM | : 220605110080 |
| Fakultas / Program Studi | : Sains dan Teknologi / Teknik Informatika |
| Judul Skripsi | : Deteksi Berita Palsu Berbahasa Indonesia Pada Data Tidak Seimbang Menggunakan *IndoBERT* Dengan Focal Loss |

Menyatakan dengan sebenarnya bahwa Skripsi yang saya tulis ini benar-benar merupakan hasil karya saya sendiri, bukan merupakan pengambil alihan data, tulisan, atau pikiran orang lain yang saya akui sebagai hasil tulisan atau pikiran saya sendiri, kecuali dengan mencantumkan sumber cuplikan pada daftar pustaka.

Apabila dikemudian hari terbukti atau dapat dibuktikan skripsi ini merupakan hasil jiplakan, maka saya bersedia menerima sanksi atas perbuatan tersebut.

Malang, 23 Juni 2026
Yang membuat pernyataan,


AKBAR BIMANTARA TRAHESTIAWAN
NIM. 220605110080



v

## MOTTO

*"Mengapa harus menjadi normal ketika Anda bisa menjadi yang terbaik?"*

(Zlatan Ibrahimovic)

## HALAMAN PERSEMBAHAN

*Alhamdulillahi rabbil 'alamin*, segala puji bagi Allah Swt. atas segala rahmat, nikmat, dan karunia-Nya. Shalawat serta salam semoga senantiasa tercurah kepada Nabi Muhammad saw., beserta keluarga, sahabat, dan seluruh umatnya. Atas izin dan pertolongan Allah Swt., penulis diberikan kemudahan dan kekuatan hingga dapat menyelesaikan skripsi ini.

Dengan penuh rasa syukur, skripsi ini penulis persembahkan kepada kedua orang tua tercinta yang selalu memberikan kasih sayang, doa, dukungan, dan pengorbanan tanpa batas. Setiap langkah dan pencapaian penulis tidak lepas dari ridha, cinta, serta perjuangan mereka.

Kepada saudara-saudara tercinta dan seluruh keluarga besar, terima kasih atas segala perhatian, semangat, motivasi, dan doa yang senantiasa mengiringi perjalanan penulis dalam menempuh pendidikan hingga mencapai tahap ini.

Kepada sahabat, teman-teman, kekasih dan semua pihak yang telah membantu dan mendukung penulis selama proses penyusunan skripsi, penulis mengucapkan terima kasih yang sebesar-besarnya. Semoga segala kebaikan yang telah diberikan mendapatkan balasan yang terbaik.

# KATA PENGANTAR

*Alhamdulillāhirabbil ‘ālamīn*, segala puji bagi Allah *Subḥānahu wa Ta‘ālā*. yang telah melimpahkan rahmat, taufik, hidayah, serta karunia-Nya sehingga penulis dapat menyelesaikan skripsi yang berjudul *“Deteksi Berita Palsu Berbahasa Indonesia Pada Data Tidak Seimbang Menggunakan IndoBERT Dengan Focal Loss”* dengan baik. Shalawat serta salam semoga senantiasa tercurah kepada junjungan kita Nabi Muhammad *Shollallohu ‘Alaihi Wasallam*, beserta keluarga, sahabat, dan seluruh pengikutnya hingga akhir zaman. Skripsi ini disusun sebagai salah satu syarat untuk memperoleh gelar Sarjana Komputer pada Program Studi Teknik Informatika, Fakultas Sains dan Teknologi, Universitas Islam Negeri Maulana Malik Ibrahim Malang. Oleh karena itu, pada kesempatan ini penulis ingin menyampaikan ucapan terima kasih yang sebesar-besarnya kepada:

1. Prof. Dr. Hj. Ilfi Nurdiana, M.Si., CAHRM., CRMP., selaku Rektor Universitas Islam Negeri Maulana Malik Ibrahim Malang.
2. Dr. Agus Mulyono, M.Kes., selaku Dekan Fakultas Sains dan Teknologi Universitas Islam Negeri Maulana Malik Ibrahim Malang.
3. Supriyono, M.Kom., selaku Ketua Program Studi Teknik Informatika Universitas Islam Negeri Maulana Malik Ibrahim Malang.
4. Dr. Zainal Abidin M.Kom., selaku dosen pembimbing I dan Nur Fitriyah Ayu Tunjung Sari, M.Cs., selaku dosen pembimbing II, yang telah memberikan bimbingan, arahan, masukan, serta motivasi kepada penulis selama proses penyusunan skripsi ini. Terima kasih atas waktu, kesabaran, dan perhatian yang telah diberikan sehingga penelitian dan penulisan skripsi ini dapat terselesaikan dengan baik dan tepat waktu.

5. Dr. Ririen Kusumawati, M.Kom., selaku dosen penguji I dan Fatchurrochman, M.Kom., selaku dosen penguji II, yang telah memberikan evaluasi, saran, serta berbagai masukan yang berharga sehingga penelitian dan penulisan skripsi ini dapat tersusun dengan lebih baik. Terima kasih atas arahan dan perhatian yang diberikan dalam proses penyusunan skripsi.
6. Seluruh dosen Program Studi Teknik Informatika Universitas Islam Negeri Maulana Malik Ibrahim Malang, penulis menyampaikan ucapan terima kasih atas ilmu yang telah di berikan selama masa perkuliahan. Semoga segala ilmu dan kebaikan yang telah diberikan menjadi amal jariyah serta memberikan manfaat yang berkelanjutan bagi penulis di masa yang akan datang.
7. Ibu tercinta Yulia Sukmawati, yang senantiasa memberikan kasih sayang, dukungan, motivasi, serta pengorbanan yang tiada henti kepada penulis. Keindahan dan ketulusan doamu yang tidak pernah putus menjadi penyerta dalam setiap perjuangan penulis, sehingga berbagai proses dan tantangan dapat dilalui hingga skripsi ini terselesaikan dengan baik.
8. Ayah tercinta Ahmad Fahrudi Setiawan, yang senantiasa menjadi teladan dalam kerja keras, keteguhan, dan tanggung jawab. Terima kasih atas doa, dukungan, serta pengorbanan yang telah diberikan selama ini. Setiap nasihat dan perjuangan beliau selalu menjadi sumber motivasi bagi penulis untuk terus berusaha hingga mampu menyelesaikan skripsi ini dengan baik dan tepat waktu.
9. Adik Aurella Nasywa Trahestiawan dan Adik Adhyatma Wasesa Trahestiawan, yang selalu hadir memberikan semangat, doa, dan dukungan kepada penulis dalam setiap proses perjalanan pendidikan.

10. Taqqiyah Daaniys Shabrina Rusydiyyah atas cinta, doa, dukungan, kesabaran, dan kepercayaan yang selalu diberikan. Terima kasih telah hadir sebagai penyemangat dalam setiap langkah perjalanan akademik ini, serta menjadi salah satu alasan bagi penulis untuk terus berusaha dan tidak menyerah hingga skripsi ini dapat terselesaikan. Semoga segala kebaikan yang telah diberikan mendapatkan balasan yang terbaik.
11. Seluruh keluarga besar penulis yang tidak dapat disebutkan satu per satu, terima kasih atas doa, dukungan, perhatian, dan semangat yang telah diberikan kepada penulis selama menempuh pendidikan.
12. Ucapan terima kasih penulis sampaikan kepada keluarga besar Teknik Informatika angkatan 2022 Infinity, yang telah memberikan kebersamaan, dukungan, dan pengalaman berharga selama masa perkuliahan.

Penulis menyadari bahwa skripsi ini masih jauh dari kata sempurna dan masih terdapat berbagai keterbatasan yang memerlukan penyempurnaan di masa mendatang. Namun demikian, penulis meyakini bahwa proses belajar tidak memiliki batas akhir, sehingga setiap kritik dan saran yang membangun akan menjadi bekal berharga untuk terus berkembang. Bagi penulis, skripsi yang baik bukanlah skripsi yang sempurna, melainkan skripsi yang dapat diselesaikan dengan penuh tanggung jawab dan tepat pada waktunya. Oleh karena itu, penulis berharap skripsi ini dapat memberikan manfaat, baik bagi pengembangan ilmu pengetahuan maupun bagi pihak-pihak yang membutuhkan sebagai referensi untuk penelitian selanjutnya.

Malang, 23 Juni 2026

Penulis

## DAFTAR ISI

## DAFTAR GAMBAR

# DAFTAR TABEL

# ABSTRAK

Trahestiawan, Akbar. 2026. **Deteksi Berita Palsu Berbahasa Indonesia Pada Data Tidak Seimbang Menggunakan IndoBERT dengan Focal Loss.** Skripsi. Jurusan Teknik Informatika Fakultas Sains dan Teknologi Universitas Islam Negeri Maulana Malik Ibrahim Malang. Pembimbing: (I)Dr. Zainal Abidin M.Kom, (II) Nur Fitriyah Ayu Tunjung Sari, M.Cs

.

Kata kunci: Deteksi Berita Palsu, Focal Loss, IndoBERT, Klasifikasi Teks, Ketidakseimbangan Data

Penyebaran berita palsu (hoaks) di ruang digital Indonesia menjadi permasalahan serius yang mengancam stabilitas sosial dan kepercayaan publik. Klasifikasi berita hoaks berbahasa Indonesia berperan penting dalam menekan penyebaran informasi palsu, namun masih menghadapi tantangan berupa ketidakseimbangan distribusi data (class imbalance). Penelitian ini mengusulkan pendekatan klasifikasi berita hoaks menggunakan teknologi *Natural Language Processing* (NLP) melalui *fine-tuning* model IndoBERT-base-p1. Dataset yang digunakan adalah Indonesia False News Dataset 2025 – Turnbackhoax.id yang terdiri dari 970 sampel, dengan distribusi 698 data kategori SALAH (71,96%) dan 272 data kategori PENIPUAN (28,04%). Penelitian dilakukan menggunakan dua skenario klasifikasi biner, yaitu Cross Entropy Loss sebagai *baseline* dan *Focal Loss* ($\alpha$=0,75; $\gamma$=2,0) untuk mengatasi ketidakseimbangan data. *Focal Loss* dirancang untuk meningkatkan pembelajaran pada sampel kelas minoritas dengan mengurangi kontribusi sampel yang mudah diklasifikasikan. Hasil evaluasi menunjukkan bahwa kombinasi IndoBERT dan Focal Loss menghasilkan akurasi 98,63% dan *macro* F1-score 98,33%, lebih baik dibandingkan Cross Entropy Loss yang memperoleh akurasi 97,26% dan *macro* F1-score 96,56%. Peningkatan terbesar terjadi pada kelas PENIPUAN dengan F1-score meningkat dari 95,00% menjadi 97,62%, sementara nilai *loss* menurun dari 0,0923 menjadi 0,0059 (93,61%). Hasil penelitian menunjukkan bahwa penerapan Focal Loss pada IndoBERT efektif meningkatkan kinerja klasifikasi berita hoaks, pada kelas minoritas.

# ABSTRACT

Trahestiawan, Akbar. 2026. **Detection of Indonesian Fake News on Imbalanced Data Using IndoBERT with Focal Loss.** Undergraduate Thesis. Department of Informatics Engineering, Faculty of Science and Technology, State Islamic University of Maulana Malik Ibrahim Malang. Advisors: (I) Dr. Zainal Abidin, M.Kom, (II) Nur Fitriyah Ayu Tunjung Sari, M.Cs.

.

Keywords: Fake News Detection, Focal Loss, IndoBERT, Text Classification, Data Imbalance

The spread of fake news (hoaxes) in Indonesia's digital space has become a serious problem that threatens social stability and public trust. The classification of Indonesian-language hoax news plays an important role in suppressing the dissemination of false information, yet it still faces challenges in the form of imbalanced data distribution (class imbalance). This study proposes a hoax news classification approach using Natural Language Processing (NLP) technology through fine-tuning of the IndoBERT-base-p1 model. The dataset used is the Indonesia False News Dataset 2025 – Turnbackhoax.id, consisting of 970 samples, with a distribution of 698 data points in the SALAH (FALSE) category (71.96%) and 272 data points in the PENIPUAN (DECEPTION) category (28.04%). The research was conducted using two binary classification scenarios: Cross Entropy Loss as a baseline and Focal Loss ($\alpha$=0.75; $\gamma$=2.0) to address data imbalance. Focal Loss is designed to enhance learning on minority class samples by reducing the contribution of easily classified samples. The evaluation results show that the combination of IndoBERT and Focal Loss achieved an accuracy of 98.63% and a macro F1-score of 98.33%, outperforming Cross Entropy Loss, which obtained an accuracy of 97.26% and a macro F1-score of 96.56%. The most significant improvement occurred in the PENIPUAN (DECEPTION) class, with the F1-score increasing from 95.00% to 97.62%, while the loss value decreased from 0.0923 to 0.0059 (93.61%). The findings indicate that the application of Focal Loss to IndoBERT is effective in improving the performance of hoax news classification, particularly for the minority class.

# مستخلص البحث


تراهستياوان، أكبر. 2026. **كشف الأخبار الكاذبة الإندونيسية في البيانا ت غير المتوازنة باستخدام نموذج IndoBERT مع دالة الخسارة .Focal Loss** رسالة البكالوريوس. قسم هندسة المعلوماتية، كلية العلوم والتكنولوجيا، جامعة مولانا مالك إبراهيم الإسلامية الحكومية مالانغ. المشرف الأول: الدكتور زين العابدين، الماجستير. المشرفة الثانية: نور فطرية آيو تونجونغ ساري، الماجستير.

الكلمات المفتاحية :كشف الأخبار الكاذبة، Focal Loss، IndoBERT، تصنيف النصوص، عدم توازن البيانات.

يُعد انتشار الأخبار الكاذبة (Hoax) في الفضاء الرقمي الإندونيسي مشكلةً خطيرةً تهدد الاستقرار الاجتماعي وثقة المجتمع. ويؤدي تصنيف الأخبار الكاذبة المكتوبة باللغة الإندونيسية دورًا مهمًا في الحد من انتشار المعلومات المضللة، إلا أنه لا يزال يواجه تحديًا يتمثل في عدم توازن توزيع البيانات .(*Data Imbalance*) تهدف هذه الدراسة إلى اقتراح منهج لتصنيف الأخبار الكاذبة باستخدام تقنيات معالجة اللغات الطبيعية (*Natural Language Processing - NLP*) من خلال Fine-Tuning لنموذج IndoBERT-base-p1 استُخدمت في هذه الدراسة مجموعة بيانات Indonesia False News Dataset Turnbackhoax.id – 2025، والتي تضم 970 عينة، تتوزع إلى 698 عينة ضمن فئة SALAH (FALSE)بنسبة 71.96%، و 272 عينة ضمن فئة PENIPUAN (DECEPTION) بنسبة 28.04% ونُفذت الدراسة باستخدام سيناريوهين للتصنيف الثنائي، وهما استخدام Cross Entropy Loss كنموذج مرجعي(*Baseline*) ، واستخدام Focal Loss بقيم $\alpha = 0.75$و $\gamma = 2.0$ لمعالجة مشكلة عدم توازن البيانات. وقد صُممت Focal Loss لتعزيز عملية التعلم على عينات الفئة الأقل تمثيلًا من خلال تقليل تأثير العينات التي يسهل تصنيفها.أظهرت نتائج التقييم أن الجمع بين نموذج IndoBERT و Focal Lossحقق دقة (Accuracy) بلغت 98.63%، ودرجة Macro F1-scoreبلغت 98.33%، متفوقًا على نموذج Cross Entropy Lossالذي حقق دقة 97.26%ودرجة Macro F1-scoreبلغت .96.56%وقد ظهر أكبر تحسن في فئة PENIPUAN (DECEPTION)، حيث ارتفعت قيمة F1-scoreمن 95.00%إلى 97.62%، في حين انخفضت قيمة Loss من 0.0923إلى 0.0059، بنسبة انخفاض بلغت .93.61% وتشير نتائج هذه الدراسة إلى أن تطبيق Focal Lossعلى نموذج IndoBERT يُعد فعالًا في تحسين أداء تصنيف الأخبار الكاذبة، ولا سيما في تصنيف عينات الفئة الأقل تمثيلًا.

# BAB I
# PENDAHULUAN

## 1.1 Latar Belakang

Salah satu tantangan terberat di ruang digital Indonesia saat ini adalah menjamurnya konten hoaks. Laporan dari Kementerian Komunikasi dan Digital (Komdigi) mencatat adanya 1.923 konten hoaks sepanjang tahun 2024, dengan puncaknya terjadi pada bulan Oktober yang mencapai 215 konten hoaks. Fenomena ini menunjukkan betapa cepatnya informasi palsu dapat tersebar dan memengaruhi opini publik, sehingga menimbulkan risiko sosial yang nyata.

Dalam perspektif Islam, permasalahan penyebaran informasi palsu ini memiliki dimensi moral dan spiritual yang mendalam. Al-Qur'an telah memberikan pedoman tegas mengenai pentingnya verifikasi informasi sebelum menyebarkannya, Sebagaimana dijelaskan di dalam Surah Al-Hujurat ayat 6.

يَٰٓأَيُّهَا ٱلَّذِينَ ءَامَنُوٓاْ إِن جَآءَكُمْ فَاسِقٌۢ بِنَبَإٖ فَتَبَيَّنُوٓاْ أَن تُصِيبُواْ قَوْمَۢا بِجَهَٰلَةٖ فَتُصْبِحُواْ عَلَىٰ مَا فَعَلْتُمْ نَٰدِمِينَ ٦

*"Wahai orang-orang yang beriman, jika seorang fasik datang kepadamu membawa berita penting, maka telitilah kebenarannya agar kamu tidak mencelakakan suatu kaum karena ketidaktahuan(mu) yang berakibat kamu menyesali perbuatanmu itu."* (Qs. Al-Hujurat[49]:6)

Prinsip etika komunikasi dalam Islam menekan pentingnya kehati-hatia dalam menerima dan menyebarkan informasi. Dalam konteks modern, konsep tabayyun dipahami sebagai kewajiban moral untuk melakukan verifikasi dan klarifikasi terhadap suatu berita sebelum diyakini atau disebarkan. Kajian kontemporer menunjukkan bahwa prinsip ini memiliki relevansi kuat dengan

tantangan penyebaran informasi di era digital. Penelitian terbaru menegaskan bahwa penafsiran fatabayyanu dalam Tafsir Al-Misbah memberikan landasan etis yang dapat diadaptasi sebagai pedoman literasi digital untuk memerangi hoaks dan disinformasi di media social (Akbar dkk., 2025). Temuan lainnya juga menggarisbawahi bahwa nilai-nilai tabayyun berperan penting dalam membentuk perilaku komunikatif yang bertanggung jawab, kritis, dan tidak tergesa-gesa dalam menerima informasi di ruang digital (Rona dkk., 2024) Analisis dataset Turnback Hoax mengungkap karakteristik khusus misinformasi di Indonesia. Dari 987 sampel, distribusi kategori menunjukkan ketidakseimbangan ekstrem: 70.7% diklasifikasikan sebagai SALAH dan 27.6% sebagai PENIPUAN, sementara kategori lain seperti "Belum Terbukti", "Satir", dan "Parodi" hanya mencakup kurang dari 2%. Pola dominasi dua kategori utama ini sejalan dengan laporan MAFINDO dan kajian misinformasi Indonesia, serta merupakan karakteristik umum dataset hoaks Indonesia yang juga dilaporkan dalam penelitian NLP terbaru (Abidin dkk., 2025)

Pembedaan antara misinformasi (informasi keliru tanpa niat menyesatkan) dan disinformasi (informasi palsu dengan niat dan motif tertentu) memiliki implikasi praktis signifikan dalam penanganan hoaks (Subekti dkk., 2025). Masing-masing memerlukan pendekatan berbeda. Misinformasi lebih efektif ditangani melalui edukasi, sementara disinformasi membutuhkan intervensi hukum dan mekanisme pelaporan serius (Kusumaningrum dkk., 2025). Namun, penelitian yang secara khusus mengkaji dan membedakan kedua kategori ini dalam konteks bahasa

Indonesia masih sangat terbatas, dengan sebagian besar studi berfokus pada hoaks secara umum tanpa membedakan dimensi intensionalitasnya.

Perkembangan *Natural Language Processing* (NLP) menawarkan solusi teknis melalui model berbasis transformer seperti IndoBERT, yang telah terbukti meningkatkan akurasi signifikan pada tugas klasifikasi bahasa Indonesia termasuk deteksi misinformasi (Koto dkk., 2021). Keunggulan ini berasal dari kemampuan attention mechanism dalam menangkap konteks semantik secara akurat.

Namun, penerapan IndoBERT pada dataset dengan ketidakseimbangan kelas ekstrem seperti kasus hoaks Indonesia menghadapi tantangan metodologis. Model tetap berisiko mengalami bias terhadap kelas mayoritas tanpa strategi penanganan ketidakseimbangan yang sistematis. Penelitian menunjukkan teknik seperti Focal Loss dapat meningkatkan kemampuan model dalam mendeteksi kelas minoritas pada klasifikasi berita hoaks berbahasa Indonesia (Abidin dkk., 2025)

## 1.2 Rumusan Masalah

Rumusan masalah dalam penelitian ini adalah:

a. Bagaimana kinerja model IndoBERT dalam klasifikasi biner konten hoaks kategori salah dan penipuan?

b. Sejauh mana teknik penanganan *class imbalance* dapat meningkatkan performa model IndoBERT, khususnya dalam mendeteksi kelas minoritas pada dataset hoaks kategori salah dan penipuan.

## 1.3 Batasan Masalah

Penelitian ini dibatasi agar fokus pada tujuan yang telah ditetapkan sebagai berikut:

a. Penelitian hanya berfokus pada klasifikasi biner untuk membedakan dua kategori hoaks yang paling dominan berdasarkan data empiris Indonesia, yaitu SALAH (yang mewakili misinformation/informasi keliru) dan PENIPUAN (yang mewakili disinformation/informasi palsu berunsur penipuan).
b. Model yang digunakan adalah IndoBERT-base-p1 dalam keadaan pretrained tanpa modifikasi arsitektur transformer dasar.
c. Dataset utama adalah "Indonesia False News Dataset 2025 - Turnbackhoax.id" yang tersedia di platform Kaggle. Dataset ini berisi konten hoaks yang telah diverifikasi dan dilabeli secara manual oleh tim Turnback Hoax. Penelitian hanya menggunakan sampel yang memiliki label salah dan penipuan sesuai dengan ruang lingkup klasifikasi biner.

## 1.4 Tujuan Penelitian

Berdasarkan rumusan masalah yang telah ditetapkan, tujuan penelitian ini adalah:

a. Mengimplementasikan model IndoBERT untuk klasifikasi biner konten hoaks pada kategori SALAH dan PENIPUAN menggunakan fungsi loss Cross Entropy sebagai model baseline.

b. Menerapkan Focal Loss pada model IndoBERT untuk menangani ketidakseimbangan data pada klasifikasi biner konten hoaks kategori SALAH dan PENIPUAN.

## 1.5 Manfaat Penelitian

Manfaat Akademis Berdasarkan rumusan masalah yang telah ditetapkan, tujuan penelitian ini adalah:

a. Penelitian ini dapat memberikan kontribusi ilmiah dalam pengembangan kajian Natural Language Processing (NLP), khususnya dalam penerapan model IndoBERT untuk klasifikasi konten hoaks.

b. penelitian ini dapat menjadi referensi bagi penelitian selanjutnya yang berkaitan dengan deteksi hoaks, penanganan ketidakseimbangan data (class imbalance), serta pengembangan model klasifikasi berbasis deep learning.

# BAB II
# STUDI PUSTAKA

Pada bab ini berisi studi pustaka yang membahas beberapa penelitian terkait yang dilakukan sebelumnya, landasan teori, dan metode yang digunakan penulis sebagai acuan mengerjakan penelitian tugas akhir ini.

## 2.1 Penelitian Terkait

IndoBERT merupakan model bahasa berbasis transformer yang dioptimalkan khusus untuk bahasa Indonesia dan telah terbukti efektif dalam berbagai tugas pemrosesan bahasa alami, termasuk analisis sentimen dan deteksi berita hoaks. Sebuah penelitian menerapkan model IndoBERT untuk mendeteksi berita hoaks berbahasa Indonesia dengan membandingkannya terhadap metode tradisional seperti SVM dan Naïve Bayes. Data penelitian diperoleh dari portal berita nasional serta unggahan media sosial relevan, yang kemudian melalui tahapan pra-pemrosesan seperti case folding, tokenisasi, stopword removal, dan stemming. Hasil pengujian menunjukkan bahwa IndoBERT unggul dengan nilai akurasi 90,14% dan F1-score 89,6%, mengungguli SVM. Keunggulan ini dikaitkan dengan kemampuan model dalam memahami konteks antar kata dalam kalimat panjang, serta adaptasinya terhadap variasi bahasa informal di media sosial (Putra dkk., t.t.)

Pada domain yang berbeda, IndoBERT juga dimanfaatkan untuk analisis sentimen pada ulasan pengguna aplikasi kesehatan di Indonesia. Fokus penelitian ini adalah kemampuan model dalam memahami ekspresi emosional pasien yang

ditulis secara tidak baku. Setelah melalui pra-pemrosesan yang mencakup case folding, tokenisasi, stopword removal, dan padding, model berhasil mencapai akurasi 96%, precision 95%, dan recall 96%. Hasil ini menunjukkan stabilitas performa model pada berbagai kategori sentimen, sekaligus membuktikan bahwa representasi kontekstual IndoBERT mampu menangkap makna semantik dan sintaktik dengan baik (Dai dkk., 2023)

Sementara itu, sebuah penelitian lain mengembangkan sistem deteksi berita palsu dengan menggabungkan IndoBERT dan teknik penyeimbangan SMOTE untuk mengatasi ketidakseimbangan data. Dataset yang digunakan dikumpulkan dari portal berita daring dan diklasifikasikan ke dalam kategori "hoaks" dan "non-hoaks". Setelah melalui tahapan pra-pemrosesan standar, model berhasil mencapai akurasi 94,3% dan F1-score 92,8%, dengan peningkatan signifikan setelah penerapan SMOTE. Temuan ini memperkuat bukti bahwa pendekatan transfer learning berbasis IndoBERT efektif dalam meningkatkan keandalan sistem deteksi berita palsu berbahasa Indonesia (Ridho & Yulianti, 2024)

Dalam upaya peningkatan performa lebih lanjut, diperkenalkan pula model hibrida yang menggabungkan IndoBERT dengan LSTM untuk klasifikasi berita hoaks. Kombinasi ini memanfaatkan kemampuan IndoBERT dalam memahami konteks semantik dan kekuatan LSTM dalam menangkap hubungan sekuensial antar token. Berdasarkan hasil evaluasi, model hibrida mencapai akurasi 93,2% dengan F1-score 92,5%, lebih tinggi dibandingkan model IndoBERT murni. Integrasi LSTM terbukti membantu memperkaya representasi temporal dan

meningkatkan kemampuan model dalam mengenali teks ambigu (Yefferson dkk., 2024)

Penelitian lain mengombinasikan metode BERTopic dan IndoBERT untuk klasifikasi berita palsu berbahasa Indonesia, dengan tujuan mengidentifikasi topik utama yang sering muncul dalam berita palsu menggunakan pendekatan clustering berbasis representasi semantik kontekstual. Hasil evaluasi menunjukkan akurasi 91,7% dengan F1-score 92,1%, membuktikan efektivitas pendekatan ini dalam mengenali pola bahasa provokatif dan informatif pada berita palsu (Prisscilya & Girsang, 2024).

Selanjutnya, sebuah studi melakukan benchmarking terhadap performa berbagai model transformer, termasuk IndoBERT, untuk klasifikasi sentimen ulasan layanan e-government berbahasa Indonesia. Hasil penelitian menunjukkan bahwa IndoBERT unggul dibandingkan model multilingual seperti mBERT dan XLM-R, dengan akurasi mencapai 94%. Keunggulan ini dikaitkan dengan representasi khusus bahasa Indonesia pada IndoBERT yang berkontribusi terhadap performa lebih stabil dan akurat pada teks layanan publik yang cenderung formal (Dhendra & Gayuh Utomo, 2025)

Penelitian terbaru menerapkan IndoBERT untuk menganalisis sentimen publik terhadap penundaan pengangkatan CPNS 2025 di platform media sosial X (Twitter). Meskipun detail tahap pra-pemrosesan tidak dijelaskan secara lengkap, model yang dilatih menggunakan dataset komentar publik berhasil menunjukkan

akurasi di atas 90%, menegaskan konsistensi performa IndoBERT dalam konteks wacana sosial digital yang dinamis (Asshiddiq & Witanti, 2025)

**Tabel 2.1 Metrik Penelitian Terdahulu**

| Penulis | Preprocessing | | | | | Metode | | | | Post-Processing / Evaluasi |
|---|---|---|---|---|---|---|---|---|---|---|
| | Case Folding | Tokenisasi | Stopword Removal | Stemming | Padding | IndoBERT | Fine Tuning | CNN / LSTM | BERTopic | Evaluasi / Post Method |
| Romadhony Dkk. (2025) | ✓ | ✓ | ✓ | ✓ | | ✓ | ✓ | | | Akurasi, F1-Score |
| Ridho & Yulianti (2025) | ✓ | ✓ | ✓ | | ✓ | ✓ | ✓ | | | Akurasi, Presisi, Recall |
| Imaduddin (2023) | ✓ | ✓ | ✓ | ✓ | | ✓ | ✓ | | | Akurasi, F1-Score, Confusion Matrix |
| Yefferson Dkk. (2024) | ✓ | ✓ | ✓ | | ✓ | ✓ | ✓ | ✓ | | Akurasi, Presisi, Recall |
| Prisscilya & Girsang (2024) | ✓ | ✓ | ✓ | ✓ | | ✓ | ✓ | | ✓ | Akurasi, F1-Score, Analisis Topik |
| Dhendra & Gayuh Utomo (2025) | ✓ | ✓ | ✓ | | | ✓ | ✓ | | | Akurasi, Presisi, Recall, F1-Score |
| Asshiddiq & Witanti (2025) | ✓ | ✓ | | | | ✓ | ✓ | | | Akurasi,Presisi,Recall,F1-Score |
| Penelitian yang dilakukan | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | | | Akurasi, Presisi, Recall, F1- Score, Support |

Berdasarkan Tabel 2.1, penelitian ini dilakukan untuk menyempurnakan penerapan model IndoBERT yang telah digunakan pada berbagai studi sebelumnya dalam deteksi berita hoaks. Sebagian besar penelitian terdahulu menunjukkan performa yang baik, namun masih terbatas pada penerapan *fine-tuning* standar dengan fungsi *loss* konvensional seperti *cross entropy*, serta belum

secara optimal menangani permasalahan ketidakseimbangan data (*class imbalance*) yang umum terjadi pada dataset hoaks.

Selain itu, tahapan *preprocessing* yang digunakan pada penelitian sebelumnya umumnya masih bersifat konvensional dan cenderung agresif, seperti penghapusan *stopword* dan *stemming* secara menyeluruh, tanpa mempertimbangkan bahwa model berbasis *Transformer* seperti IndoBERT memiliki kemampuan memahami konteks bahasa secara mendalam. Pendekatan tersebut berpotensi menghilangkan informasi linguistik penting yang justru diperlukan dalam membedakan karakteristik antara kategori SALAH dan PENIPUAN.

Berbeda dengan penelitian sebelumnya, penelitian ini tidak mengombinasikan IndoBERT dengan model lain seperti CNN atau LSTM, melainkan berfokus pada optimalisasi performa IndoBERT melalui dua pendekatan utama. Pertama, penerapan *preprocessing minimal* yang selektif untuk mempertahankan struktur linguistik penting dalam teks. Kedua, penggunaan fungsi *loss* berbasis *Focal Loss* sebagai pengganti *cross entropy* untuk mengatasi permasalahan *class imbalance* dengan memberikan fokus lebih besar pada data minoritas.

Selain itu, dilakukan penyesuaian parameter pelatihan (*fine-tuning*) pada IndoBERT, seperti *learning rate*, jumlah *epoch*, serta penggunaan *optimizer* AdamW untuk meningkatkan stabilitas dan kecepatan konvergensi model. Dengan pendekatan ini, diharapkan model yang dihasilkan menjadi lebih sensitif terhadap konteks bahasa, lebih akurat dalam melakukan klasifikasi, serta mampu

memberikan performa yang konsisten pada dataset berita berbahasa Indonesia dengan distribusi kelas yang tidak seimbang.

## 2.2 Pemrosesan Bahasa Alami (*Natural Language Processing*)

Natural Language Processing (NLP) merupakan cabang dari kecerdasan buatan (*Artificial Intelligence*) yang berfokus pada kemampuan komputer untuk memahami, mengolah, dan menganalisis bahasa manusia dalam bentuk teks maupun suara. NLP menggabungkan konsep dari ilmu linguistik, pembelajaran mesin, dan ilmu komputer untuk memungkinkan sistem memahami makna, konteks, serta struktur bahasa secara otomatis. Dalam perkembangan terbaru, NLP telah menjadi komponen penting dalam berbagai aplikasi seperti analisis sentimen, klasifikasi teks, penerjemahan bahasa, hingga deteksi berita palsu. Kemampuan NLP dalam memahami hubungan antar kata dan konteks kalimat menjadikannya sangat efektif dalam mengolah data teks dalam jumlah besar. Selain itu, pendekatan modern dalam NLP telah beralih dari metode berbasis aturan ke metode berbasis *deep learning* yang mampu menghasilkan representasi bahasa yang lebih kompleks dan akurat (Hu dkk., 2025)

Proses dalam NLP umumnya melibatkan beberapa tahapan penting, yaitu *preprocessing*, representasi teks, dan pemodelan. Tahap *preprocessing* mencakup pembersihan data seperti *case folding*, tokenisasi, penghapusan *stopword*, dan stemming atau lemmatization. Setelah itu, teks diubah menjadi bentuk numerik melalui teknik representasi seperti *Bag of Words*, *TF-IDF*, atau *word embedding*. Pada tahap pemodelan, data yang telah direpresentasikan kemudian digunakan sebagai input untuk algoritma *machine learning* atau *deep learning* guna

melakukan tugas tertentu, seperti klasifikasi teks. Dalam konteks deteksi berita palsu, tahapan ini sangat penting untuk memastikan bahwa model dapat memahami pola bahasa yang membedakan antara berita hoaks dan non-hoaks. Oleh karena itu, kualitas proses NLP sangat berpengaruh terhadap performa model yang dihasilkan (Aslan dkk., 2024)

Perkembangan NLP saat ini didominasi oleh penggunaan model berbasis *transformer* yang mampu menangkap konteks bahasa secara lebih mendalam dibandingkan metode sebelumnya. Model seperti BERT dan variannya, termasuk IndoBERT, menggunakan mekanisme *self-attention* untuk memahami hubungan antar kata dalam suatu kalimat secara dua arah (*bidirectional*). Pendekatan ini terbukti lebih efektif dalam berbagai tugas NLP, termasuk klasifikasi teks dan deteksi berita palsu. Selain itu, model berbasis *deep learning* juga mampu menangani kompleksitas bahasa alami yang seringkali ambigu dan kontekstual. Dalam penelitian ini, NLP berperan sebagai fondasi utama dalam proses pengolahan data teks berita sebelum dilakukan klasifikasi menggunakan model IndoBERT dengan Focal Loss. Dengan demikian, penerapan NLP yang tepat diharapkan dapat meningkatkan akurasi dan efektivitas sistem dalam mendeteksi berita palsu pada data berbahasa Indonesia (Roumeliotis dkk., 2025)

### 2.3 Berita Palsu

Berita palsu atau yang lebih dikenal dengan istilah hoaks merupakan informasi yang tidak sesuai dengan kebenaran atau keadaan yang sebenarnya. Secara etimologis, kata hoaks berasal dari bahasa Inggris hoax yang dalam Kamus Besar Bahasa Indonesia (KBBI) diartikan sebagai berita bohong, sedangkan

Oxford Dictionary mendefinisikan hoax sebagai "a humorous or malicious deception" yang berarti tipuan yang bersifat lucu atau berniat jahat (Oxford University Press, 2023). Menurut Dedi Rianto Rahadi, berita palsu didefinisikan sebagai usaha untuk menipu atau mengakali pembaca agar mempercayai sesuatu, padahal pencipta berita tersebut mengetahui bahwa berita itu palsu (Rahadi, 2017). Berita palsu juga merupakan pemberitaan palsu yang merupakan upaya pemutarbalikan fakta menggunakan informasi yang seolah-olah meyakinkan tetapi tidak dapat diverifikasi kebenarannya, serta informasi yang direkayasa untuk menutupi informasi yang sebenarnya (Zainuddin, 2018). Hoaks atau berita palsu memiliki tujuan yang beragam, mulai dari sekadar lelucon atau iseng, propaganda politik untuk menjatuhkan citra lawan politik (black campaign), penipuan ekonomi melalui promosi dengan modus penipuan, hingga agitasi untuk menghasut atau mempengaruhi reaksi emosional masyarakat (Wardhani, 2019).

Dalam perspektif information disorder atau kekacauan informasi, berita palsu dapat dikategorikan menjadi tiga jenis berdasarkan unsur kesengajaannya (Wardle & Derakhshan, 2017). Pertama, misinformasi yaitu informasi yang salah atau tidak akurat, tetapi disebarkan tanpa niat jahat karena penyebar percaya bahwa informasi tersebut benar. Kedua, disinformasi yaitu informasi palsu yang sengaja dibuat dan disebarluaskan untuk menipu, memengaruhi opini publik, atau mengaburkan kebenaran. Ketiga, malinformasi yaitu informasi yang benar tetapi disajikan secara keluar dari konteks untuk merugikan pihak tertentu. Dengan demikian, berita palsu dalam pengertian luas mencakup baik misinformasi

maupun disinformasi, di mana perbedaan utamanya terletak pada ada tidaknya unsur kesengajaan di balik penyebaran informasi tersebut.

Berita palsu memiliki perbedaan fundamental dengan berita penipuan, meskipun keduanya sering kali dianggap sama dalam praktik sehari-hari. Berita palsu secara umum adalah informasi bohong yang disebarkan tanpa atau dengan niat tertentu, sedangkan berita penipuan merupakan subset dari disinformasi yang secara spesifik bertujuan untuk merugikan korban secara materiil, terutama dalam konteks transaksi elektronik (Kominfo, 2023). Dengan kata lain, berita penipuan adalah disinformasi dengan motif ekonomi yang sengaja dirancang untuk menguntungkan pelaku secara finansial dengan merugikan korban, mengandung unsur penyesatan konsumen, dan disebarkan secara terencana (Wardani, 2019). Berbeda dengan misinformasi yang terjadi tanpa niat, berita penipuan merupakan tindakan yang direncanakan dan memiliki dampak kerugian material bagi korbannya.

Berdasarkan modus operasi yang teridentifikasi di lapangan, berita penipuan dapat berbentuk beberapa jenis. Pertama, konten tiruan (imposter content) yaitu berita yang mengatasnamakan pejabat, lembaga, atau tokoh publik untuk meminta uang atau donasi. Kedua, konten palsu (fabricated content) yaitu dokumen atau surat palsu yang mengatasnamakan instansi resmi. Ketiga, tautan atau tagihan palsu berupa file APK atau tautan phishing yang mengatasnamakan PLN, BPJS, atau kurir (MAFINDO, 2024). Sementara itu, berita palsu kategori misinformasi yang tidak mengandung unsur penipuan biasanya berbentuk klaim hoaks tentang bencana alam, video lama yang dikaitkan dengan peristiwa baru, atau informasi

kesehatan yang tidak terverifikasi (Turnbackhoax, 2023). Perbedaan mendasar antara kedua kategori ini terletak pada ada tidaknya unsur kesengajaan dan motif ekonomi dari pelaku penyebar informasi.

Secara hukum positif Indonesia, penyebaran berita bohong diatur dalam beberapa pasal. Pasal 28 ayat (1) Undang-Undang Informasi dan Transaksi Elektronik (UU ITE) menyatakan bahwa "Setiap Orang dengan sengaja, dan tanpa hak menyebarkan berita bohong dan menyesatkan yang mengakibatkan kerugian konsumen dalam Transaksi Elektronik" dapat dikenai sanksi pidana. Selain itu, Pasal 390 Kitab Undang-Undang Hukum Pidana (KUHP) mengatur tentang penyebaran kabar bohong untuk menguntungkan diri sendiri atau orang lain, sedangkan Pasal 14 Undang-Undang Nomor 1 Tahun 1946 mengatur tentang penyebaran berita bohong yang dapat menerbitkan keonaran di kalangan rakyat (Munir, 2020). Perbedaan pengaturan hukum ini mencerminkan bahwa berita palsu dan berita penipuan memiliki konsekuensi hukum yang berbeda sesuai dengan tingkat kesengajaan dan dampak yang ditimbulkan.

Dalam konteks penelitian ini, kategori SALAH dan PENIPUAN yang digunakan sebagai objek klasifikasi mencerminkan dua jenis berita palsu yang berbeda. Kategori SALAH merepresentasikan misinformasi, yaitu informasi yang tidak benar namun disebarkan tanpa niat jahat, seringkali karena kelalaian verifikasi. Kategori ini mencakup klaim hoaks tentang bencana alam, video lama yang dikaitkan dengan peristiwa baru, atau informasi kesehatan yang tidak terverifikasi. Sementara itu, kategori PENIPUAN merepresentasikan disinformasi dengan motif ekonomi, yaitu informasi bohong yang sengaja dibuat untuk

merugikan korban secara materiil, seperti akun tiruan pejabat yang meminta uang, surat palsu mengatasnamakan instansi, atau tautan phishing mengatasnamakan layanan publik (Abidin dkk., 2025). Pembedaan ini menjadi dasar penting dalam penelitian karena kedua kategori memerlukan pendekatan penanganan yang berbeda. Misinformasi lebih efektif ditangani melalui edukasi dan literasi digital, sementara disinformasi membutuhkan intervensi hukum dan mekanisme pelaporan yang lebih serius (Kusumaningrum dkk., 2025). Dengan memahami perbedaan ini, sistem deteksi hoaks yang dikembangkan dalam penelitian ini diharapkan dapat memberikan klasifikasi yang lebih akurat dan kontekstual sesuai dengan karakteristik masing-masing kategori berita palsu.

### 2.4 Representasi Teks dan Word Embedding

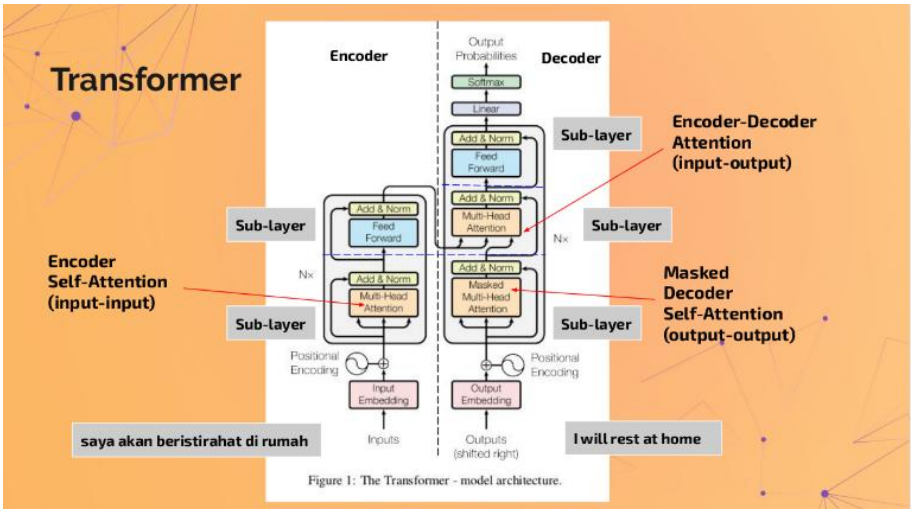
Representasi teks adalah proses mengonversi data linguistik menjadi bentuk numerik yang dapat diproses oleh model pembelajaran mesin. Dalam model berbasis *transformer* seperti *IndoBERT*, representasi diperoleh melalui *embedding kontekstual* yang dilatih menggunakan *Masked Language Modeling (MLM)*. Pada *MLM*, sejumlah *token* dalam kalimat disembunyikan, dan model dilatih untuk menebak *token* yang hilang berdasarkan konteks kata di sekitarnya. Model mempelajari hubungan antar *token* secara simultan dari kiri dan kanan *token* yang diprediksi. Setiap *token* dalam kalimat dipetakan ke vektor berdimensi tetap, yang memuat informasi kontekstual dari seluruh kalimat. *Token [CLS]* merepresentasikan keseluruhan kalimat atau dokumen, sedangkan *token* lain menghasilkan *embedding kontekstual* yang dapat dianalisis untuk memahami makna kata dan frasa. *Embedding* ini berbeda dari metode klasik seperti

*Word2Vec* atau *GloVe* karena memperhitungkan konteks seluruh kalimat, bukan hanya hubungan kata secara lokal. Proses *embedding* dalam *IndoBERT* dimulai dengan *tokenisasi* teks menjadi *sub-word*, diikuti dengan penambahan *positional encoding* untuk menyandikan urutan *token*. Vektor *token* kemudian diproses melalui lapisan *encoder* yang terdiri dari *multi-head self-attention* dan *feed-forward network*. Setiap lapisan *encoder* menghasilkan representasi *token* yang diperbarui berdasarkan hubungan dengan *token* lain dalam kalimat, menghasilkan *embedding kontekstual* untuk setiap *token* dan *token [CLS]*. Vektor-vektor ini menjadi input lapisan klasifikasi untuk tugas tertentu, termasuk klasifikasi teks seperti deteksi *fake news*. Penelitian menunjukkan bahwa model berbasis *IndoBERT* mampu mencapai akurasi tinggi dalam klasifikasi berita palsu berbahasa Indonesia. Dataset yang lebih besar dan beragam meningkatkan kemampuan model mengenali pola baru. Analisis pola bahasa emosional dan hiperbolis dapat membantu model membedakan *hoaks* dari berita faktual. *Embedding* kata dan kalimat merepresentasikan makna dan konteks teks secara efektif. Pelatihan model dilakukan secara iteratif untuk mengurangi kesalahan prediksi. Penyesuaian model terhadap konteks budaya lokal meningkatkan relevansi prediksi. Sistem pembelajaran mesin dapat diterapkan dalam platform digital untuk deteksi berita palsu secara otomatis (Lin dkk., t.t.).

### 2.5 Arsitektur Transformer

Arsitektur *Transformer* diperkenalkan oleh Vaswani et al. untuk memproses urutan teks secara paralel dengan memanfaatkan mekanisme *self-attention*, yang memungkinkan setiap *token* dalam kalimat mempelajari hubungan dengan semua

*token* lain dalam urutan yang sama, sehingga konteks kata dapat direpresentasikan secara menyeluruh. *Transformer* terdiri dari dua komponen utama, yaitu *Encoder* dan *Decoder*, di mana *Encoder* membangun representasi kontekstual dari input, sedangkan *Decoder* digunakan pada tugas generatif, seperti penerjemahan atau prediksi *token* berikutnya.




**Gambar 2. 1 Alur kerja arsitektur Transformer**

(Sumber : Ruben Tea (2023)

Gambar 2.1 menunjukkan alur kerja arsitektur Transformer yang terdiri dari dua komponen utama, yaitu Encoder dan Decoder, di mana Encoder membangun representasi kontekstual dari input, sedangkan Decoder digunakan untuk tugas generative Pada tahap awal, kalimat input melalui proses *preprocessing*, di mana kata-kata diubah menjadi *token*, ditambahkan *positional encoding* untuk menyandikan urutan kata, dan dikonversi menjadi representasi numerik. *Token* ini kemudian dimasukkan ke dalam *Encoder Transformer*, yang memproses informasi melalui lapisan *Self-Attention* untuk menghitung bobot perhatian antar *token*. Contohnya, kata “beristirahat” memperhatikan kata “saya” dalam konteks

kalimat, sehingga dihasilkan *contextual embeddings*. Representasi ini kemudian diperkuat melalui *Feed Forward Layer* dan normalisasi (*Add & Norm*), membentuk vektor kontekstual untuk setiap *token*. Setiap lapisan *Encoder* memperbarui representasi *token* berdasarkan hubungan dengan *token* lain dalam kalimat, menghasilkan embedding kontekstual yang kaya informasi. Vektor-vektor ini menjadi input untuk lapisan klasifikasi pada tugas seperti deteksi *fake news*.

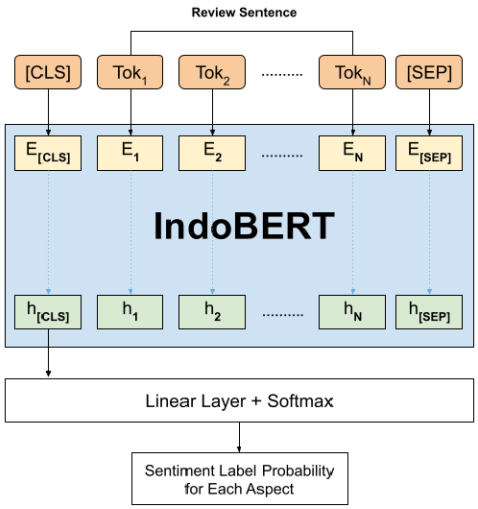
### 2.6 Arsitektur *IndoBERT*

IndoBERT merupakan model *pre-trained language model* berbasis arsitektur *Bidirectional Encoder Representations from Transformers* (BERT) yang secara khusus dikembangkan untuk memahami bahasa Indonesia. Model ini dilatih menggunakan korpus teks berbahasa Indonesia dalam jumlah besar sehingga mampu menangkap karakteristik linguistik, struktur kalimat, serta konteks semantik yang khas dalam bahasa Indonesia. IndoBERT memanfaatkan pendekatan *deep learning* untuk menghasilkan representasi teks yang lebih kontekstual dibandingkan metode tradisional seperti *Bag of Words* atau *TF-IDF*. Dengan kemampuan memahami konteks secara mendalam, IndoBERT banyak digunakan dalam berbagai tugas NLP seperti klasifikasi teks, analisis sentimen, dan deteksi berita palsu. Keunggulan utama IndoBERT terletak pada kemampuannya dalam memahami hubungan antar kata dalam suatu kalimat secara lebih akurat (Koto dkk., 2021)

Secara arsitektur, IndoBERT mengadopsi model Transformer yang menggunakan mekanisme *self-attention* untuk memproses hubungan antar kata

dalam suatu urutan teks. Mekanisme ini memungkinkan model untuk memperhatikan setiap kata dalam kalimat secara bersamaan dan menentukan tingkat kepentingannya terhadap kata lainnya. Selain itu, IndoBERT bersifat *bidirectional*, yang berarti model membaca konteks dari dua arah sekaligus, yaitu dari kiri ke kanan dan dari kanan ke kiri. Hal ini berbeda dengan model sebelumnya seperti LSTM yang memproses teks secara sekuensial. Dengan pendekatan ini, IndoBERT mampu menghasilkan pemahaman konteks yang lebih komprehensif dan meningkatkan performa dalam berbagai tugas klasifikasi teks. Kemampuan ini menjadikan IndoBERT sebagai salah satu model yang efektif dalam menangani kompleksitas bahasa alami. (Devlin dkk., t.t.)

Dalam implementasinya, IndoBERT biasanya digunakan melalui proses *fine-tuning*, yaitu melatih kembali model yang telah *pre-trained* menggunakan dataset spesifik sesuai dengan tugas yang diinginkan. Pada penelitian ini, IndoBERT digunakan untuk melakukan klasifikasi berita menjadi hoaks dan non-hoaks berdasarkan teks yang dianalisis. Proses *fine-tuning* memungkinkan model untuk menyesuaikan bobotnya dengan karakteristik data penelitian sehingga dapat meningkatkan akurasi prediksi. Selain itu, IndoBERT juga mampu bekerja dengan baik pada dataset berukuran besar maupun kecil karena telah memiliki pengetahuan awal dari proses *pre-training*. Dengan demikian, penggunaan IndoBERT dalam penelitian ini diharapkan dapat meningkatkan performa sistem dalam mendeteksi berita palsu berbahasa Indonesia secara lebih akurat dan efektif (Koto dkk., 2021)




**Gambar 2. 2 Arsitektur *IndoBERT* untuk klasifikasi text (Sumber : Evi Yulianti (2024))**

Gambar 2.2 di atas menunjukkan arsitektur IndoBERT yang telah di-fine-tune untuk tugas klasifikasi teks. Input berupa kalimat (misalnya, ulasan produk) di-tokenisasi menjadi token-token sub-kata. Token-token ini kemudian diproses melalui model IndoBERT untuk menghasilkan representasi kontekstual. Representasi token [CLS] digunakan sebagai representasi keseluruhan kalimat dan diteruskan ke lapisan linear diikuti dengan fungsi aktivasi softmax untuk menghasilkan probabilitas kelas.

## 2.7 *Fine Tuning* IndoBERT

Pada tahap *fine-tuning* IndoBERT, model dilatih kembali untuk menyesuaikan parameter internalnya dengan dataset yang digunakan dalam penelitian. Proses ini bertujuan agar model yang telah melalui tahap *pre-training* dapat memahami karakteristik khusus dari data penelitian, seperti pola bahasa dalam berita hoaks dan non-hoaks. Dalam proses pelatihan ini digunakan

algoritma AdamW (*Adaptive Moment Estimation with Weight Decay*), yang merupakan pengembangan dari optimizer Adam dengan penambahan mekanisme *weight decay* untuk mengurangi risiko *overfitting*. AdamW bekerja dengan mengombinasikan momentum pertama dan momentum kedua untuk meningkatkan stabilitas dan kecepatan konvergensi selama pelatihan model.

Langkah awal dalam algoritma AdamW adalah menghitung momentum pertama dan momentum kedua berdasarkan gradien hasil proses *backpropagation*. Momentum pertama merupakan rata-rata eksponensial dari gradien, sedangkan momentum kedua merupakan rata-rata eksponensial dari kuadrat gradien. Perhitungan keduanya dirumuskan sebagai berikut:

$$m_t = \beta_1 m_{t-1} + (1 - \beta_1) g_t \tag{2.1}$$

$$v_t = \beta_2 v_{t-1} + (1 - \beta_2) g_t^2 \tag{2.2}$$

Keterangan :
$m_t$ adalah estimasi momentum pertama pada iterasi ke-$t$
$v_t$ adalah estimasi momentum kedua pada iterasi ke- $t$
$g_t$ adalah gradien pada iterasi ke-$t$
$\beta_1$ dan $\beta_2$ merupakan parameter decay yang biasanya bernilai 0.9 dan 0.999.

Setelah itu, dilakukan proses *bias correction* untuk mengoreksi nilai momentum yang masih bias pada iterasi awal pelatihan. Hal ini diperlukan karena pada tahap awal nilai $m_t$ dan $v_t$ cenderung mendekati nol sehingga belum representatif. Proses koreksi bias dilakukan dengan persamaan berikut:

$$\widehat{m_t} = \frac{m_t}{1 - \beta_1^t}, \qquad \widehat{v_t} = \frac{v_t}{1 - \beta_2^t} \tag{2.3}$$

Keterangan :
$m_t$ adalah estimasi momentum pertama setelah koreksi bias
$v_t$ adalah estimasi momentum kedua setelah koreksi bias

Selanjutnya, parameter model diperbarui menggunakan rumus AdamW yang telah memperhitungkan learning rate, momentum yang telah dikoreksi, serta weight decay. Persamaan pembaruan parameter dituliskan sebagai berikut:

$$\theta_{t+1} = \theta_t - \eta \cdot \frac{\widehat{m_t}}{\sqrt{\widehat{v_t}} + \epsilon} - \eta \cdot \lambda \cdot \theta_t \quad (2.4)$$

Keterangan :
$m_t$ adalah estimasi momentum pertama setelah koreksi bias
$v_t$ adalah estimasi momentum kedua setelah koreksi bias

Dalam proses *fine-tuning*, setelah parameter diperbarui, dilakukan tahap *forward pass* di mana data teks yang telah melalui tokenisasi diubah menjadi representasi numerik berupa *logits*. Nilai logits ini kemudian dikonversi menjadi probabilitas menggunakan fungsi *Softmax,* yang berfungsi untuk memetakan skor ke dalam distribusi probabilitas. Fungsi Softmax dirumuskan sebagai berikut:

$$\mathrm{P}(\mathrm{y_i}) = \frac{\mathrm{e}^{\mathrm{z_i}}}{\sum_{\mathrm{j=1}}^{\mathrm{K}} \mathrm{e}^{\mathrm{z_j}}} \quad (2.5)$$

Keterangan :
$P(y_i)$ adalah probabilitas untuk kelas ke-$i$
$z_i$ adalah nilai logits pada kelas ke$-i$
$K$adalah jumlah kelas
$e$ adalah bilangan eksponensial

Melalui fungsi ini, model dapat menentukan kelas dengan probabilitas tertinggi sebagai hasil prediksi akhir. Setelah diperoleh probabilitas dari fungsi Softmax, langkah selanjutnya adalah menghitung nilai $loss$ menggunakan fungsi Cross Entropy. Fungsi ini digunakan untuk mengukur perbedaan antara distribusi probabilitas hasil prediksi model dengan label sebenarnya. Rumus Cross Entropy dituliskan sebagai berikut:

$$L = -\sum_{i=1}^{K} y_i \log(p_i) \quad (2.6)$$

Keterangan :
$L$ adalah nilai loss
$y_i$ adalah label sebenarnya (one-hot encoding)
$p_i$ adalah probabilitas hasil prediksi dari Softmax
$K$ adalah jumlah kelas

Dalam kasus klasifikasi biner, rumus Cross Entropy dapat disederhanakan menjadi:

$$L = -(y \log(p) + (1 - y) \log(1 - p)) \quad (2.7)$$

Fungsi ini akan memberikan penalti yang lebih besar apabila model memberikan probabilitas rendah pada kelas yang sebenarnya benar.

**2.8 Focal Loss**

Focal Loss merupakan fungsi *loss* yang dikembangkan untuk mengatasi permasalahan *class imbalance* dalam proses pelatihan model *machine learning* dan *deep learning*. Pada dataset yang tidak seimbang, model cenderung bias terhadap kelas mayoritas sehingga mengabaikan kelas minoritas yang justru sering menjadi fokus utama, seperti pada kasus deteksi berita palsu. Focal Loss bekerja dengan cara mengurangi kontribusi dari sampel yang mudah diklasifikasikan dan memberikan bobot lebih besar pada sampel yang sulit (*hard examples*). Dengan demikian, model akan lebih fokus belajar pada data yang sulit dan meningkatkan kemampuan dalam mendeteksi kelas minoritas. Pendekatan ini terbukti efektif dalam berbagai tugas klasifikasi yang memiliki distribusi data tidak seimbang (Lin dkk., t.t.)

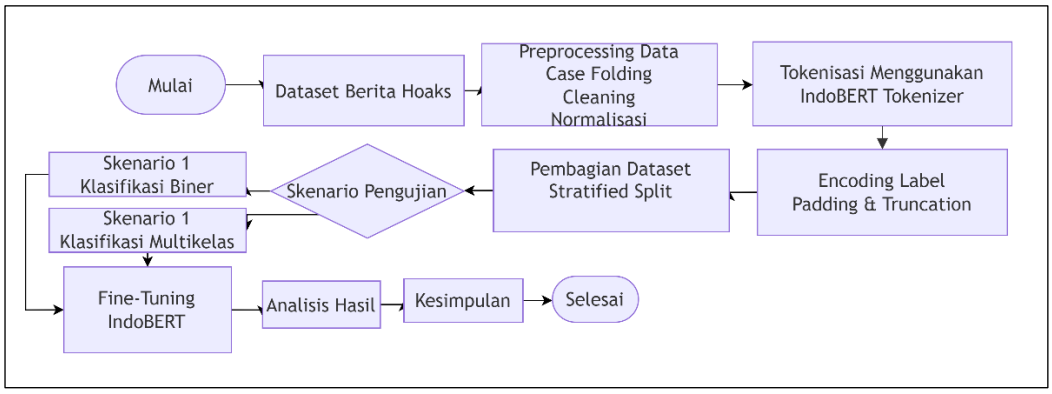
Secara matematis, Focal Loss merupakan pengembangan dari *Cross Entropy Loss* dengan penambahan faktor modulasi yang dikontrol oleh parameter gamma (γ). Parameter ini berfungsi untuk mengatur tingkat fokus terhadap sampel yang sulit diklasifikasikan. Semakin besar nilai gamma, maka semakin kecil kontribusi dari sampel yang mudah terhadap nilai loss. Selain itu, Focal Loss juga dapat dikombinasikan dengan parameter alpha (α) untuk memberikan bobot yang berbeda pada masing-masing kelas, sehingga membantu mengatasi ketidakseimbangan distribusi data. Dengan adanya dua parameter tersebut, Focal Loss menjadi fleksibel dalam menyesuaikan kebutuhan model terhadap kondisi dataset yang digunakan.

Dalam implementasinya, Focal Loss banyak digunakan pada berbagai bidang seperti deteksi objek, klasifikasi citra, serta klasifikasi teks berbasis NLP. Pada penelitian ini, Focal Loss digunakan untuk meningkatkan performa model IndoBERT dalam mendeteksi berita palsu pada dataset yang tidak seimbang. Penggunaan Focal Loss diharapkan dapat meningkatkan nilai *recall* dan *F1-score* pada kelas minoritas (hoaks), sehingga model tidak hanya memiliki akurasi tinggi tetapi juga mampu mengenali berita palsu secara lebih efektif. Dengan demikian, kombinasi antara IndoBERT dan Focal Loss menjadi pendekatan yang tepat dalam mengatasi tantangan klasifikasi teks pada data yang tidak seimbang (Beseiso & Al-Zahrani, 2025)

# BAB III
# METODOLOGI PENELITIAN

## 3.1 Rancangan Penelitian

Tahapan penelitian ini mengikuti alur eksperimental untuk mengembangkan dan mengevaluasi model deteksi hoaks berbasis IndoBERT dengan fokus pada penanganan ketidakseimbangan data. Proses dimulai dari identifikasi masalah dan pengumpulan dataset Turnbackhoax.id, dilanjutkan dengan preprocessing dan pembagian data. Tahap kritis melibatkan penerapan dua teknik penanganan ketidakseimbangan, Baseline dan Focal Loss. Selanjutnya dilakukan tokenisasi dan fine-tuning model IndoBERT. Model dievaluasi dengan fokus pada recall, dan hasilnya dianalisis secara komparatif untuk menarik kesimpulan mengenai efektivitas teknik yang diuji. Gambar 3.1 menunjukkan Tahapan Penelitian,




**Gambar 3. 1 Diagram Prosedur Penelitian**

## 3.2 Dataset

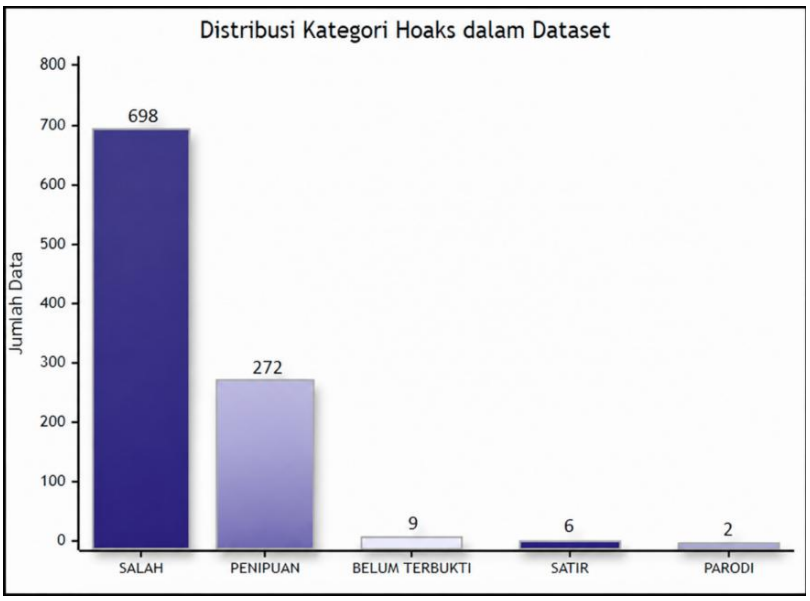
Sumber data yang digunakan dalam penelitian ini berasal dari *public dataset* yang tersedia pada platform Kaggle dengan judul *Indonesia False News Dataset 2025 – TurnBackHoax.id*. *Dataset* ini memuat berita dalam bahasa Indonesia yang

telah diklasifikasikan per kategori. Format data disajikan dalam bentuk *table*, di mana setiap baris merepresentasikan satu berita dengan informasi terkait tanggal publikasi, judul, isi berita, dan *label* kebenaran. Atribut utama dalam *dataset* meliputi *id*, *date*, *title*, *content*, *label*, dan *source*. Penjelasan masing-masing fitur ditunjukkan pada Tabel 3.1 Deskripsi Data.

**Tabel 3. 1 Deskripsi Data**

| Fitur | Penjelasan |
|---|---|
| *Id* | Nomor unik setiap berita |
| *Date* | Tanggal publikasi berita |
| *Title* | Judul berita |
| *Content* | Isi berita secara lengkap |
| *Flag* | Kategori berita |
| *Source* | Sumber berita |

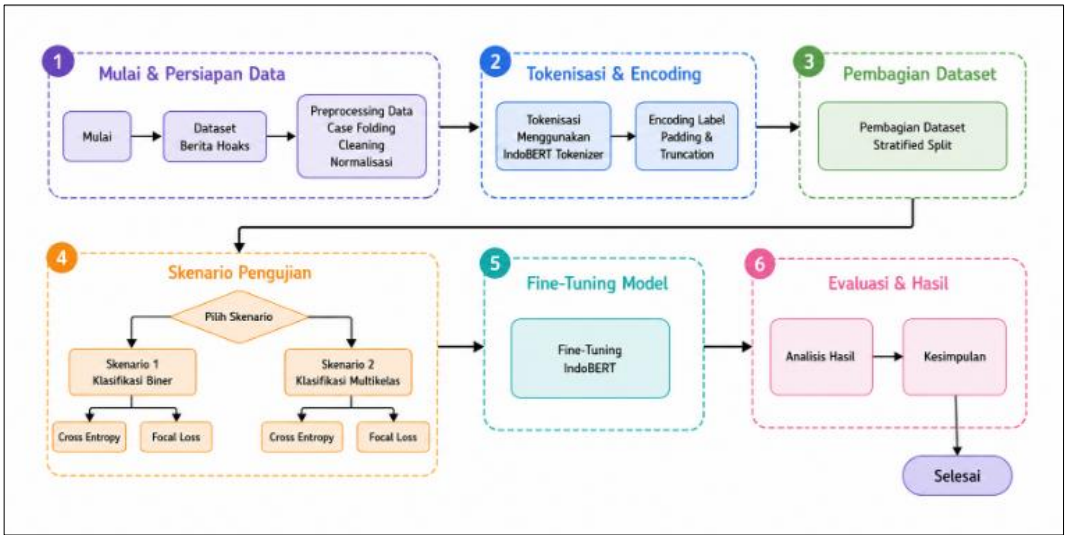
Dataset TurnBackHoax.id mencakup berita yang diterbitkan antara tahun 2019 hingga 2025, sehingga mampu mewakili variasi topik dan gaya bahasa di media sosial dan portal berita Indonesia. Jumlah total data sekitar 900 entri, dengan data mayoritas diisi dengan label SALAH, lalu yang kedua adalah label PENIPUAN dengan jumlah lebih dari 250 data, dan label yang lain seperti BELUM, SATIR, PARODI yang mempunyai jumlah kurang dari 10% dari keseluruhan data. Gambar 3.2 menunjukkan distribusi kategori yang ada pada dataset.




**Gambar 3. 2 Distribusi Kategori Dataset**

### 3.3 Desain Sistem

Desain sistem penelitian ini menggambarkan alur pemodelan deteksi hoaks berbasis IndoBERT. Sistem dimulai dari pengumpulan dataset Turnbackhoax.id, dilanjutkan dengan preprocessing teks dan pembagian data. Tahap kritis melibatkan penerapan dua teknik penanganan ketidakseimbangan data: Baseline dan Focal Loss. Selanjutnya dilakukan tokenisasi IndoBERT, pelatihan model, dan evaluasi performa. Hasil dari kedua teknik kemudian dibandingkan untuk dianalisis efektivitasnya, diakhiri dengan kesimpulan penelitian. Gambar 3.3 menunjukkan desain sistem penelitian ini.




**Gambar 3. 3 Desain Model**

### 3.4 Filtering Data SALAH dan PENIPUAN

Tahap preprocessing dilakukan untuk menyiapkan teks berita dari dataset Turnbackhoax.id agar dapat diolah secara optimal oleh model *IndoBERT*. Proses ini bertujuan untuk membersihkan, menormalkan, dan menyeragamkan data teks sehingga menghasilkan representasi yang konsisten dan mudah dipahami oleh algoritma pembelajaran mesin. Setiap tahap preprocessing saling berhubungan untuk memastikan kualitas data yang digunakan pada proses pelatihan dan pengujian model (Faisal & Mahendra, 2022).

Tahap awal dalam penelitian ini adalah melakukan filtering data untuk menyesuaikan dataset dengan tujuan penelitian, yaitu klasifikasi berita hoaks ke dalam tiga kategori. Dataset asli terdiri dari beberapa kategori hoaks yang digunakan oleh TurnBackHoax. Dalam penelitian ini, kategori yang digunakan adalah SALAH, PENIPUAN, dan OTHERS. Kategori OTHERS dibentuk dengan

menggabungkan beberapa kategori yang memiliki jumlah data relatif sedikit, yaitu BELUM TERBUKTI, SATIR, dan PARODI.

Proses filtering dilakukan dengan mempertahankan seluruh data yang berlabel SALAH dan PENIPUAN, kemudian mengelompokkan data dari kategori BELUM TERBUKTI, SATIR, dan PARODI ke dalam satu kategori baru, yaitu OTHERS. Pendekatan ini bertujuan untuk mengurangi permasalahan jumlah sampel yang sangat sedikit pada masing-masing kategori minor sekaligus mempertahankan informasi yang terkandung dalam data tersebut. Dengan demikian, model dapat melakukan klasifikasi berita hoaks ke dalam tiga kelas yang berbeda, yaitu SALAH, PENIPUAN, dan OTHERS.

Hasil filtering menghasilkan dataset sebanyak 987 sampel, yang terdiri dari 698 data SALAH (70,72%), 272 data PENIPUAN (27,56%), dan 17 data OTHERS (1,72%). Distribusi data menunjukkan adanya ketidakseimbangan kelas (class imbalance), terutama pada kategori OTHERS yang memiliki jumlah sampel jauh lebih sedikit dibandingkan kategori lainnya. Oleh karena itu, penelitian ini menerapkan teknik penanganan ketidakseimbangan data pada tahap pelatihan model untuk meningkatkan kemampuan klasifikasi pada seluruh kelas.

### 3.5 Data Prapemrosesan

*Preprocessing data* merupakan tahap penting untuk mempersiapkan teks berita agar siap digunakan dalam proses pelatihan model deteksi hoaks berbasis IndoBERT. Tujuannya adalah meningkatkan kualitas dan konsistensi data dengan membersihkan teks dari elemen yang tidak relevan, sehingga model dapat belajar

pola bahasa Indonesia secara optimal. Berbeda dengan pendekatan preprocessing konvensional yang sering kali agresif dalam menghilangkan fitur teks, penelitian ini menerapkan *preprocessing minimal* yang selektif untuk mempertahankan informasi linguistik penting yang dapat menjadi sinyal pembeda antara kategori hoaks. Pendekatan ini sejalan dengan penelitian terbaru yang menunjukkan bahwa model berbasis *Transformer* seperti BERT dan turunannya (termasuk IndoBERT) lebih efektif ketika menggunakan teks yang relatif utuh karena telah memiliki kemampuan memahami konteks secara mendalam melalui mekanisme *self-attention* (Sanh dkk., 2020).

Selain itu, beberapa studi terkini menunjukkan bahwa preprocessing yang terlalu agresif, seperti penghapusan *stopword* dan *stemming* secara berlebihan, justru dapat menghilangkan informasi semantik penting yang dibutuhkan oleh model *Transformer* dalam memahami konteks kalimat (Gao dkk., 2022). Oleh karena itu, penelitian ini menggunakan pendekatan *preprocessing minimal* yang hanya berfokus pada pembersihan *noise* seperti karakter khusus, URL, dan simbol yang tidak relevan, tanpa menghilangkan struktur linguistik utama dari teks.

Tahap preprocessing dalam penelitian ini meliputi beberapa langkah utama yang disesuaikan dengan karakteristik model IndoBERT dan fokus penelitian pada klasifikasi biner dan multikelas, sehingga data yang dihasilkan tetap informatif sekaligus optimal untuk proses pelatihan model.

a. *Cleansing*

Proses cleansing dilakukan secara selektif dengan fokus pada penghapusan elemen yang benar-benar mengganggu dan tidak informatif. Elemen yang dihapus meliputi URL, mention (@username), dan hashtag (#), sementara tanda baca penting seperti titik (.), koma (,), tanda tanya (?), dan tanda seru (!) dipertahankan karena dapat memberikan informasi tentang struktur kalimat dan nada penulisan. Angka dan karakter khusus yang relevan dengan konteks berita juga tidak dihilangkan. Proses cleansing ini bertujuan untuk membersihkan noise tanpa mengurangi kandungan informasi semantik yang esensial.

**Tabel 3. 1 Cleansing**

| Teks Sebelum | Teks Sesudah |
|---|---|
| Id "Breaking News!!! Vaksin Covid-19 menyebabkan mutasi genetik pada manusia. Ini adalah konspirasi global untuk mengontrol populasi dunia!!! *#Hoax #FakeNews"* | Breaking News Vaksin Covid menyebabkan mutasi genetik pada manusia Ini adalah konspirasi global untuk mengontrol populasi dunia |

b. *Case Folding*

*Case folding* adalah proses mengubah semua huruf dalam teks menjadi huruf kecil untuk memastikan keseragaman. Tabel 3.4 dibawah merupakan contoh *case Folding.*

**Tabel 3. 2 Case Folding**

| Teks Sebelum | Teks Sesudah |
|---|---|
| Breaking News Vaksin Covid menyebabkan mutasi genetik pada manusia Ini adalah konspirasi global untuk mengontrol populasi dunia | breaking news vaksin covid menyebabkan mutasi genetik pada manusia ini adalah konspirasi global untuk mengontrol populasi dunia |

*c. Stopword Removal*

Stopword removal adalah proses menghilangkan kata-kata umum yang sering muncul tetapi memiliki nilai informasi yang rendah, seperti "yang", "di", "ke", "dari", "dan", "atau". Penelitian ini menggunakan daftar stopword bahasa Indonesia yang berisi 759 kata (Tala, 2003). Proses ini membantu mengurangi dimensi data dan memfokuskan model pada kata-kata yang lebih informatif. Namun, beberapa stopword yang mungkin penting untuk konteks tertentu dipertahankan berdasarkan analisis preliminary.

*d. Stemming*

Stemming adalah proses mengembalikan kata ke bentuk dasarnya dengan menghilangkan afiks (prefiks, sufiks, infiks). Penelitian ini menggunakan algoritma Sastrawi, stemmer khusus untuk bahasa Indonesia yang dikembangkan oleh Fakultas Ilmu Komputer Universitas Indonesia. Contoh: kata "menggunakan" menjadi "guna", "pemerintah" menjadi "perintah". Stemming membantu mengurangi variasi morfologis dan memperkaya kemunculan kata dasar dalam korpus.

### 3.6 Split Dataset

Setelah semua tahap preprocessing selesai, dataset dibagi menjadi tiga bagian menggunakan teknik stratified split untuk memastikan distribusi kelas yang proporsional di setiap subset. Pembagian dilakukan dengan rasio 70:15:15, yaitu 70% untuk data training, 15% untuk data validation, dan 15% untuk data

testing. Stratifikasi memastikan bahwa persentase sampel dari setiap kategori (SALAH , PENIPUAN dan OTHERS) tetap sama di ketiga subset, sehingga model dapat dilatih, divalidasi, dan diuji pada distribusi data yang representatif. Pendekatan ini sangat penting dalam konteks ketidakseimbangan kelas untuk memastikan evaluasi model yang valid dan generalisasi yang baik.

**Tabel 3. 3 Pembagian stratified split**

| Subset | Persentase | Jumlah Data (SALAH) | Jumlah Data (PENIPUAN) | Jumlah Data (Others) | Total |
|---|---|---|---|---|---|
| Training | 70% | 489 | 190 | 12 | 679 |
| Validation | 15% | 105 | 41 | 3 | 148 |
| Testing | 15% | 104 | 41 | 2 | 148 |
| Total | 100% | 698 | 272 | 17 | 987 |

Pembagian dataset ini memungkinkan pelatihan model yang efektif dengan data yang cukup, validasi yang tepat selama training untuk mencegah overfitting, dan evaluasi akhir yang andal pada data yang belum pernah dilihat oleh model.

### 3.7 Tokenisasi dan Encoding *IndoBERT*

Setelah data teks dibersihkan melalui tahap preprocessing, langkah selanjutnya adalah tokenisasi dan encoding menggunakan tokenizer dari model IndoBERT. Tujuan dari tahap ini adalah mengubah teks mentah menjadi representasi numerik yang dapat diproses oleh model BERT. Pada tahapan ini, proses tokenisasi menggunakan *WordPiece Tokenization* yang merupakan mekanisme bawaan BERT untuk memecah kata menjadi sub-kata. Metode ini berguna untuk menangani kata-kata yang jaknrang muncul (*OOV – Out of Vocabulary*) (Devlin dkk., t.t.). Proses tokenisasi dilakukan melalui beberapa langkah utama. Pertama, ditambahkan token khusus *[CLS]* di awal dan *[SEP]* di

akhir kalimat. Kedua, setiap kata dipecah menjadi sub-kata dengan prefix ## untuk menunjukkan lanjutan dari kata sebelumnya. Ketiga, setiap token diubah menjadi *token ID* sesuai *vocabulary* IndoBERT. Keempat, dilakukan *padding* hingga panjang maksimal (*max length*) 512 token agar semua input memiliki ukuran yang sama saat masuk ke model. Tabel 3.4 merupakan contoh tokenisasi indobert.

**Tabel 3. 4 Tokenisasi *IndoBERT***

| Langkah | Hasil |
|---|---|
| **Teks asli** | Breaking News!!! Vaksin Covid-19 menyebabkan mutasi genetik pada manusia. Ini adalah konspirasi global untuk mengontrol populasi dunia!!! |
| **Tokenisasi** | ['breaking', 'news', '!', '!', '!', 'vaksin', 'covid', '-', '19', 'menyebabkan', 'mutasi', 'genetik', 'pada', 'manusia', '.', 'ini', 'adalah', 'konspirasi', 'global', 'untuk', 'mengontrol', 'populasi', 'dunia', '!', '!', '!'] |
| **Token Khusus** | ['[CLS]', 'breaking', 'news', '!', '!', '!', 'vaksin', 'covid', '-', '19', 'menyebabkan', 'mutasi', 'genetik', 'pada', 'manusia', '.', 'ini', 'adalah', 'konspirasi', 'global', 'untuk', 'mengontrol', 'populasi', 'dunia', '!', '!', '!', '[SEP]'] |
| **Konversi Token IDs** | [2, 9881, 55, 4425, 30457, 30457, 30457, 11087, 2108, 7427, 30469, 426, 1618, 11949, 11447, 126, 666, 30470, 92, 154, 22052, 4092, 90, 8361, 6169, 594, 30457, 30457, 30457, 3 |
| **Padding** | [[2, 9881, 55, 4425, 30457, 30457, 30457, 11087, 2108, 7427, 30469, 426, 1618, 11949, 11447, 126, 666, 30470, 92, 154, 22052, 4092, 90, 8361, 6169, 594, 30457, 30457, 30457, 3, 0, 0] |
| **Attention Mask** | [1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 0, 0] |

Data ini siap digunakan sebagai input ke model IndoBERT untuk proses pelatihan klasifikasi berita hoaks.

### 3.8 Training *IndoBERT*

Pelatihan model klasifikasi dengan IndoBERT dilakukan menggunakan arsitektur Transformer yang menerapkan mekanisme self-attention untuk memahami konteks kata secara bidirectional. Model IndoBERT-base-p1 yang telah dilatih sebelumnya pada korpus besar bahasa Indonesia (seperti Wikipedia

Indonesia, berita, dan media sosial) diinisialisasi menggunakan fungsi AutoModelForSequenceClassification dari library Transformers. Kode inisialisasi model dapat dimuat dengan skrip yang ditunjukan pada gambar 3.3

```
from transformers import AutoModelForSequenceClassification
model = AutoModelForSequenceClassification.from_pretrained(
    "indobenchmark/indobert-base-p1",
    num_labels=2
)
```

**Gambar 3. 4 Potongan kode memanggil model dengan library transformer**

Model ini memiliki total 124,443,651 parameter yang mencakup lapisan embedding, 12 layer transformer dengan mekanisme self-attention, dan feed-forward network. Klasifikasi biner untuk kategori SALAH (0) dan PENIPUAN (1) diimplementasikan dengan menambahkan classification head pada output layer model.

Algoritma AdamW Optimizer digunakan dalam proses pelatihan dengan learning rate sebesar $2 \times 10^{-5}$. Proses pelatihan menggunakan linear learning rate warmup selama 10% dari total training steps untuk stabilisasi awal, dilanjutkan dengan linear decay. Batch size ditetapkan sebesar 16 dengan gradient accumulation sebanyak 4 langkah sehingga menghasilkan effective batch size 64. Pelatihan berjalan maksimum 10 epoch dengan mekanisme early stopping berdasarkan nilai validation loss dengan patience selama 3 epoch.

Layer-wise learning rate decay diterapkan dengan pembagian tingkat pembaruan parameter: lapisan atas (layer 11-12) menggunakan learning rate

$2 \times 10^{-5}$, lapisan tengah (layer 7-10) menggunakan $1.5 \times 10^{-5}$, dan lapisan bawah (layer 1-6) menggunakan $1 \times 10^{-5}$. Strategi ini memungkinkan lapisan yang lebih dekat dengan output beradaptasi lebih cepat terhadap tugas deteksi hoaks.

Selama proses pelatihan, model dievaluasi untuk memperoleh estimasi performa yang lebih robust. Fungsi loss yang digunakan bervariasi sesuai teknik penanganan ketidakseimbangan data: Cross-Entropy Loss untuk baseline dan Focal Loss untuk skenario penyeimbangan kelas. Gradient clipping dengan maksimum norm 1.0 digunakan untuk mencegah exploding gradient, sementara dropout sebesar 0.1 diterapkan sebagai regularisasi pada hidden states, attention probabilities, dan classifier layers.

Pelatihan menggunakan Automatic Mixed Precision (AMP) untuk mengoptimalkan penggunaan memori GPU dan mempercepat komputasi, dengan forward dan backward pass dijalankan dalam presisi FP16 sementara optimizer states disimpan dalam FP32. Random seed ditetapkan pada 42 untuk memastikan reproduktibilitas seluruh eksperimen.

### 3.9 Fine-Tuning IndoBERT

Pada tahap Fine-Tuning IndoBERT, model dilatih untuk mempelajari dan menyesuaikan parameter internalnya berdasarkan data pelatihan yang telah diproses. Pada proses ini digunakan library *BertTokenizer* yang terdapat pada framework Transformers untuk mengonversi teks menjadi token yang dapat diproses oleh model BERT, serta *BertForSequenceClassification* yang digunakan

untuk membangun model klasifikasi berbasis IndoBERT dalam mengklasifikasikan berita menjadi kategori Salah dan Penipuan.

Selain itu, digunakan algoritma AdamW, yang merupakan varian dari Adam optimizer yang menggabungkan weight decay untuk regularisasi. AdamW merupakan algoritma optimisasi yang mengombinasikan momentum pertama dan momentum kedua untuk meningkatkan kecepatan konvergensi dan stabilitas selama pelatihan. Penambahan weight decay bertujuan untuk mengurangi kompleksitas model sehingga dapat meminimalkan risiko overfitting (Loshchilov & Hutter, 2019).

Proses pembaruan parameter dalam AdamW dilakukan melalui tiga langkah utama, yaitu pembaruan momentum pertama (estimasi gradien), momentum kedua (estimasi kuadrat gradien), serta pembaruan parameter model. Agar model dapat bekerja secara optimal, proses fine-tuning dilakukan dengan parameter yang ditunjukkan pada tabel dibawah:

**Tabel 3. 5 Parameter Fine Tuning**

| **Hyperparameter** | **Nilai** |
|---|---|
| Learning Rate | 2e-5 |
| *β1* | 0.9 |
| β2 | 0.999 |
| **ε** | 1e-8 |
| Weight Decay | 0.01 |

(Sumber : Loshchilov & Hutter (2019))

Proses pembaruan parameter dimulai dengan menghitung momentum pertama dan momentum kedua berdasarkan gradien hasil backpropagation. Secara matematis, kedua momentum tersebut dirumuskan sebagai berikut:

$$m_t = \beta_1 m_{(t-1)} + (1 - \beta_1) g_t \quad (3.1)$$

$$v_t = \beta_2 v_{(t-1)} + (1-\beta_2) g_t^2 \tag{3.2}$$

Momentum pertama $m_t$ merepresentasikan arah rata-rata gradien, sedangkan momentum kedua $v_t$ merepresentasikan besarnya variasi gradien. Sebagai ilustrasi pada iterasi pertama (t = 1), dengan nilai awal $m_0$=0 , $v_0$ =0, dan gradien $g_t$=0.5, diperoleh:

$$m_1 = 0.9 \times 0 + (1-0.9) \times 0.5 = 0.05$$

$$v_1 = 0.999 \times 0 + (1-0.999) \times (0.5)^2 = 0.00025$$

Nilai ini menunjukkan bahwa model mulai menangkap arah dan skala perubahan gradien pada iterasi awal. Namun, karena nilai momentum masih bias terhadap nol, maka dilakukan koreksi bias menggunakan persamaan berikut:

$$\hat{m}_t = m_t/(1-\beta_1^t) \tag{3.3}$$

$$\hat{v}_t = v_t/(1-\beta_2^t) \tag{3.4}$$

Dengan memasukkan nilai sebelumnya, diperoleh:

$$\hat{m_1} = 0.05/(1-0.9) = 0.5$$

$$\hat{v_1} = 0.00025/(1-0.999) = 0.25$$

Proses koreksi ini bertujuan untuk menghasilkan estimasi momentum yang lebih akurat, terutama pada tahap awal pelatihan. Selanjutnya, parameter model diperbarui menggunakan rumus AdamW yang telah mempertimbangkan learning rate, momentum yang telah dikoreksi, serta weight decay:

$$\theta_{(t+1)} = \theta_t - \eta \times \left(\hat{m}_t / \left(\sqrt{(\hat{v}_t)} + \varepsilon\right)\right) - \eta \times \lambda \times \theta_t \tag{3.5}$$

Sebagai contoh, jika nilai awal parameter θ0=1\theta_0 = 1θ0=1, maka:

$$\theta_1 = 1 - 2 \times 10^{-}5 \times \left(0.5/\left(\sqrt{0.}25 + 10^{-}8\right)\right) - 2 \times 10^{-}5 \times 0.01$$

Karena $\sqrt{0.}25 = 0.5$, maka:

$$\theta_1 = 1 - 2 \times 10^{-}5 \times 1 - 2 \times 10^{-}7$$

$$\theta_1 = 0.9999798$$

Nilai ini menunjukkan bahwa parameter model mengalami pembaruan secara bertahap dengan perubahan yang kecil, sehingga proses pembelajaran berlangsung stabil dan terkontrol.

Setelah parameter diperbarui, dilakukan proses *forward pass*, di mana teks yang telah ditokenisasi diubah menjadi representasi numerik berupa logits. Sebagai ilustrasi, digunakan teks "Video Jet Tempur Rusia tiba di Biak Papua", dan diasumsikan model menghasilkan nilai logits sebesar 1.2 untuk kelas SALAH dan 2.0 untuk kelas PENIPUAN. Nilai ini kemudian dikonversi menjadi probabilitas menggunakan fungsi Softmax berikut:

$$P(y_i = c) = e^{(z_c)}/\Sigma_{(j=1)}^{C} e^{(z_j)} \quad (3.6)$$

Perhitungan dilakukan dengan mencari nilai eksponensial dari masing-masing logit, yaitu $e^{1.2} = 3.320$ dan $e^{2.0} = 7.389$ kemudian dijumlahkan menjadi 10.709. Probabilitas masing-masing kelas diperoleh dengan membagi nilai eksponensial terhadap total tersebut, sehingga didapatkan probabilitas 0.310 untuk kelas SALAH dan 0.690 untuk kelas PENIPUAN. Hasil ini menunjukkan

bahwa model memprediksi teks tersebut sebagai kategori PENIPUAN dengan tingkat keyakinan sebesar 69%.

Untuk mengukur kesalahan prediksi, digunakan fungsi Cross Entropy Loss yang dirumuskan sebagai:

$$L = -\Sigma y_c log(p_c) \tag{3.7}$$

Karena label sebenarnya adalah PENIPUAN, maka diperoleh

$$L = -log(0.690)$$

$$L = 0.371$$

Nilai loss yang relatif kecil menunjukkan bahwa prediksi model cukup mendekati label sebenarnya. Selain itu, penelitian ini juga menggunakan Focal Loss untuk mengatasi ketidakseimbangan data, yang dirumuskan sebagai:

$$L_F L = -\alpha(1 - p_t)^\gamma log(p_t) \tag{3.8}$$

Dengan parameter α=0.75, γ=2, dan $p_t$ = 0.690, diperoleh nilai loss sebesar 0.0267. Nilai ini lebih kecil dibandingkan Cross Entropy, yang menunjukkan bahwa Focal Loss mengurangi kontribusi sampel yang mudah dan memberikan fokus lebih besar pada sampel yang sulit diklasifikasikan. Setelah nilai Focal Loss dihitung, langkah selanjutnya adalah melakukan proses backpropagation untuk menghitung gradien dari setiap parameter model terhadap nilai loss tersebut. Gradien ini menunjukkan seberapa besar pengaruh setiap parameter terhadap besarnya nilai loss, sehingga parameter dengan gradien besar akan mendapatkan perubahan yang lebih signifikan dibandingkan parameter dengan gradien kecil. Proses backpropagation dilakukan secara berurutan dari lapisan output hingga

lapisan input model, di mana gradien dihitung menggunakan aturan rantai (chain rule) yang memanfaatkan turunan parsial dari fungsi loss terhadap setiap bobot dan bias pada jaringan. Gradien yang telah dihitung kemudian digunakan oleh optimizer AdamW untuk memperbarui bobot model melalui mekanisme pembaruan parameter yang telah dijelaskan pada persamaan (3.5). Dengan demikian, model secara bertahap menyesuaikan bobotnya untuk meminimalkan nilai loss, sehingga prediksi yang dihasilkan menjadi semakin akurat seiring berjalannya proses pelatihan.

Proses perhitungan loss, backpropagation, dan pembaruan bobot ini berulang untuk setiap batch data dalam satu epoch. Satu epoch didefinisikan sebagai satu siklus lengkap di mana seluruh data training telah diproses oleh model. Pada setiap epoch, model memproses data secara bertahap dalam batch-batch kecil, di mana setiap batch berisi sejumlah sampel data yang diproses secara paralel untuk mempercepat komputasi. Setelah seluruh batch dalam satu epoch selesai diproses, model dievaluasi pada data validasi untuk menghitung validation loss dan validation F1-score. Evaluasi ini bertujuan untuk memantau performa model pada data yang tidak digunakan selama pelatihan, sehingga dapat diketahui apakah model masih terus belajar atau sudah mulai mengalami overfitting. Jika validation loss terus menurun seiring bertambahnya epoch, maka model masih dalam proses pembelajaran yang baik. Namun, jika validation loss mulai meningkat meskipun training loss terus menurun, hal ini mengindikasikan bahwa model mulai overfitting terhadap data training.

Untuk mencegah terjadinya overfitting, penelitian ini menerapkan teknik early stopping yang telah disebutkan sebelumnya. Early stopping bekerja dengan memantau nilai validation loss pada setiap epoch dan menghentikan proses pelatihan jika validation loss tidak mengalami penurunan dalam beberapa epoch berturut-turut. Pada penelitian ini, patience ditetapkan sebesar 3 epoch, yang berarti jika validation loss tidak menurun selama 3 epoch berturut-turut, maka proses pelatihan akan dihentikan dan model dengan validation loss terbaik dari epoch-epoch sebelumnya akan disimpan sebagai model akhir. Pendekatan ini memastikan bahwa model yang dihasilkan tidak hanya memiliki performa yang baik pada data training, tetapi juga mampu memberikan prediksi yang akurat pada data baru yang belum pernah dilihat sebelumnya. Model dengan bobot terbaik ini kemudian disimpan dalam format file yang dapat digunakan kembali untuk proses inferensi pada tahap evaluasi.

## 3.10 Skenario Pengujian

Penelitian ini dirancang dengan dua skenario pengujian untuk menganalisis pengaruh jumlah kelas terhadap performa model IndoBERT dalam tugas klasifikasi berita hoaks berbahasa Indonesia. Kedua skenario menerapkan tahapan preprocessing, arsitektur model, parameter pelatihan, dan metrik evaluasi yang identik. Dengan demikian, perbedaan hasil performa yang diperoleh dapat diatribusikan secara langsung pada perubahan jumlah kelas dalam proses klasifikasi.

### 3.10.1 Skenario Klasifikasi Biner (2 Kelas)

Skenario pertama menerapkan pendekatan klasifikasi biner dengan dua kategori, yaitu SALAH dan PENIPUAN. Pada skenario ini, hanya data yang memiliki salah satu dari kedua label tersebut yang digunakan dalam proses pelatihan dan pengujian. Setelah dilakukan proses filtering, diperoleh total 970 data, dengan rincian 698 data berlabel SALAH dan 272 data berlabel PENIPUAN.

Tujuan utama dari skenario ini adalah untuk menetapkan performa dasar (baseline performance) model IndoBERT pada tugas klasifikasi hoaks dua kelas. Hasil pengujian dari skenario ini kemudian dijadikan tolok ukur (benchmark) untuk membandingkan performa model ketika jumlah kelas ditingkatkan pada skenario berikutnya. Hal ini penting dilakukan agar setiap peningkatan atau penurunan performa pada skenario multikelas dapat terukur secara objektif.

### 3.10.2 Skenario Klasifikasi Multikelas (3 Kelas)

Skenario kedua menggunakan pendekatan klasifikasi multikelas dengan tiga kategori, yaitu SALAH, PENIPUAN, dan OTHERS. Kategori OTHERS dibentuk dengan menggabungkan beberapa kategori minoritas yang memiliki jumlah data relatif sangat sedikit, yaitu BELUM TERBUKTI, SATIR, dan PARODI. Penggabungan ini dilakukan untuk menguji kemampuan model dalam menangani kelas tambahan tanpa mengurangi jumlah data pelatihan secara signifikan.

Setelah proses penggabungan, diperoleh total 987 data, dengan komposisi 698 data SALAH, 272 data PENIPUAN, dan 17 data OTHERS. Penambahan kategori OTHERS bertujuan untuk menguji kemampuan model IndoBERT dalam

menangani klasifikasi multikelas. Selain itu, skenario ini juga mengevaluasi sejauh mana ketidakseimbangan data (imbalanced data) yang ekstrem dapat memengaruhi stabilitas dan akurasi prediksi model.

Melalui skenario ini, dapat dianalisis apakah model IndoBERT tetap mampu mempertahankan performa yang kompetitif ketika dihadapkan pada jumlah kelas yang lebih banyak dan distribusi data yang lebih timpang. Analisis ini penting untuk mengetahui batas toleransi model terhadap kompleksitas kelas. Hasil yang diperoleh juga akan memberikan gambaran mengenai tantangan nyata dalam implementasi sistem deteksi hoaks di lapangan.

### 3.10.3 Perbandingan Antar Skenario

Perbandingan antara skenario klasifikasi biner dan multikelas dilakukan untuk menganalisis dampak penambahan kategori OTHERS terhadap performa model secara kuantitatif dan kualitatif. Evaluasi performa dilakukan dengan menggunakan metrik Akurasi, Presisi, Recall, dan F1-Score. Selain itu, Confusion Matrix juga digunakan untuk mengidentifikasi pola kesalahan klasifikasi pada masing-masing kelas.

**Tabel 3. 6 Perbandingan Skenario Pengujian**

| **Skenario** | **Jumlah Kelas** | **PENIPUAN** |
|---|---|---|
| Klasifikasi Biner | 2 | SALAH(698), PENIPUAN (272) |
| Klasifikasi Multikelas | 3 | SALAH(698),PENIPUAN(272),OTHERS (17) |

Hasil dari kedua skenario tersebut selanjutnya dibandingkan secara sistematis untuk mengetahui pengaruh penambahan kelas minoritas terhadap kemampuan generalisasi model IndoBERT. Analisis ini diharapkan dapat

memberikan wawasan mendalam mengenai tantangan klasifikasi berita hoaks pada data dengan distribusi yang tidak seimbang. Temuan dari penelitian ini juga diharapkan dapat menjadi kontribusi praktis bagi pengembangan sistem deteksi hoaks berbasis bahasa Indonesia di masa mendatang.

### 3.11 Evaluasi Model

Evaluasi model dalam penelitian ini bertujuan untuk menilai kinerja hasil fine-tuning model IndoBERT dalam mengklasifikasikan teks berita ke dalam dua kategori utama, yaitu SALAH dan PENIPUAN. Proses evaluasi dilakukan untuk memastikan bahwa model yang dikembangkan tidak hanya mampu melakukan prediksi, tetapi juga memiliki tingkat ketepatan yang tinggi dalam mengenali pola bahasa yang membedakan kedua kategori hoaks ini. Evaluasi menjadi bagian penting dalam tahap validasi hasil fine-tuning karena menentukan seberapa baik model memahami konteks semantik dari data yang diberikan, terutama dalam menangani ketidakseimbangan kelas.

Dalam penelitian ini, metode evaluasi yang digunakan meliputi Confusion Matrix biner serta empat metrik utama, yaitu akurasi (accuracy), presisi (precision), daya ingat (recall), dan F1-score. Confusion Matrix biner berbentuk tabel 2 × 2 yang menggambarkan jumlah data yang diklasifikasikan dengan benar maupun salah pada dua kategori. Melalui Confusion Matrix, peneliti dapat mengetahui secara langsung jumlah True Positive (TP), False Positive (FP), True Negative (TN), dan False Negative (FN) untuk masing-masing kelas. Nilai-nilai ini kemudian menjadi dasar dalam perhitungan metrik evaluasi seperti presisi,

recall, dan F1-score. Struktur dasar Confusion Matrix biner ditunjukkan pada Tabel 3.5.

**Tabel 3. 7 Confusion Matrix**

| **Aktual / Prediksi** | **SALAH** | **PENIPUAN** | **OTHERS** |
|---|---|---|---|
| **SALAH** | C11 | C12 | C13 |
| **PENIPUAN** | C21 | C22 | C23 |
| **OTHERS** | C31 | C32 | C33 |

Keterangan :
C11 adalah data SALAH diprediksi SALAH
C12 adalah data SALAH diprediksi PENIPUAN
C13 adalah data SALAH diprediksi OTHERS
C21 adalah data PENIPUAN diprediksi SALAH
C22 adalah data PENIPUAN diprediksi PENIPUAN
C23 adalah data PENIPUAN diprediksi OTHERS
C31 adalah data OTHERS diprediksi SALAH
C32 adalah data OTHERS diprediksi PENIPUAN
C33 adalah data OTHERS diprediksi OTHERS

Metrik evaluasi yang digunakan dalam penelitian ini terdiri dari empat ukuran utama yang saling melengkapi. Akurasi (accuracy) digunakan untuk mengukur seberapa besar persentase prediksi yang benar terhadap keseluruhan data uji. Presisi (precision) menunjukkan tingkat ketepatan model dalam memprediksi teks hoaks, yaitu seberapa banyak data yang diprediksi sebagai hoaks benar-benar merupakan hoaks. Recall (daya ingat) digunakan untuk mengetahui kemampuan model dalam mengenali seluruh data hoaks yang sebenarnya ada. F1-score merupakan rata-rata harmonik antara presisi dan recall yang digunakan untuk mengukur keseimbangan antara keduanya, terutama dalam kasus di mana data tidak seimbang antara dua kelas, dengan definisi:

$$\text{Precision} = \frac{TP}{TP + FP} \tag{3.3}$$

$$\text{Recall} = \frac{TP}{TP + FN} \tag{3.4}$$

$$F_1 = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \tag{3.5}$$

di mana $TP$, $FP$, dan $FN$ adalah jumlah true positive, false positive, dan false negative.

# BAB IV

# HASIL DAN PEMBAHASAN

## 4.1 Tahap Pengujian

Tahap pengujian merupakan proses untuk mengevaluasi kinerja model IndoBERT dalam melakukan klasifikasi berita hoaks berbahasa Indonesia. Model yang digunakan dalam penelitian ini adalah IndoBERT Base (indobenchmark/indobert-base-p1) yang telah melalui proses fine-tuning sesuai dengan tahapan yang dijelaskan pada Bab III. Pada tahap ini, model diuji menggunakan data uji (test set) yang terpisah dari data latih untuk memastikan bahwa model tidak hanya mampu menghafal data pelatihan, tetapi juga memiliki kemampuan generalisasi yang baik terhadap data baru.

Pengujian dilakukan untuk mengetahui sejauh mana model mampu mempelajari pola-pola yang terdapat dalam data latih serta mengaplikasikan pengetahuan tersebut dalam proses klasifikasi pada data yang belum pernah dilihat sebelumnya. Kemampuan generalisasi ini menjadi aspek penting karena menentukan efektivitas model ketika diterapkan pada kasus nyata dalam mendeteksi berbagai kategori berita hoaks.

Sesuai dengan skenario pengujian yang telah dijelaskan pada Bab III, penelitian ini menggunakan dua skenario klasifikasi, yaitu klasifikasi biner dan klasifikasi multikelas. Pada skenario klasifikasi biner, model melakukan klasifikasi ke dalam dua kategori, yaitu SALAH dan PENIPUAN. Sementara itu,

pada skenario klasifikasi multikelas, model melakukan klasifikasi ke dalam tiga kategori, yaitu SALAH, PENIPUAN, dan OTHERS.

Selain membandingkan jumlah kelas, penelitian ini juga membandingkan performa model berdasarkan dua fungsi loss, yaitu Cross Entropy Loss sebagai model baseline dan Focal Loss sebagai pendekatan untuk menangani ketidakseimbangan kelas. Perbandingan dilakukan untuk menganalisis pengaruh fungsi loss terhadap kemampuan model dalam melakukan klasifikasi pada dataset dengan distribusi data yang tidak seimbang.

Evaluasi performa model dilakukan menggunakan beberapa metrik, yaitu Accuracy, Precision, Recall, dan F1-Score. Selain itu, digunakan pula Confusion Matrix untuk melihat distribusi prediksi yang benar maupun kesalahan klasifikasi pada setiap kategori sehingga dapat diketahui pola kesalahan yang dihasilkan model pada masing-masing skenario pengujian.

### 4.1.1 Data Pengujian

Dataset yang digunakan dalam penelitian ini telah melalui tahapan preprocessing sebagaimana dijelaskan pada Bab III, yang meliputi pembersihan teks, normalisasi, tokenisasi menggunakan tokenizer IndoBERT, serta proses encoding label ke dalam bentuk numerik. Tahapan tersebut bertujuan untuk memastikan bahwa data telah siap digunakan dalam proses pelatihan dan pengujian model berbasis transformer.

Penelitian ini menggunakan dua skenario pengujian, yaitu klasifikasi biner dan klasifikasi multikelas. Pada skenario klasifikasi biner, dataset terdiri dari dua

label, yaitu SALAH dan PENIPUAN dengan total 970 data. Sementara itu, pada skenario klasifikasi multikelas, dataset terdiri dari tiga label, yaitu SALAH, PENIPUAN, dan OTHERS dengan total 987 data.

Pembagian dataset dilakukan menggunakan metode stratified split dengan rasio 70:15:15 untuk data training, validation, dan testing. Penggunaan stratified split bertujuan untuk mempertahankan proporsi setiap kelas pada seluruh subset sehingga distribusi data tetap representatif. Distribusi data uji pada skenario klasifikasi multikelas ditunjukkan sebagai berikut:

**Tabel 4. 1 Jumlah data training dan testing tiap rasio**

| Kelas | Jumlah |
|---|---|
| SALAH | 105 |
| PENIPUAN | 41 |
| OTHERS | 3 |
| TOTAL | 149 |

Distribusi data menunjukkan adanya ketidakseimbangan kelas yang cukup tinggi, terutama pada kategori OTHERS yang hanya memiliki tiga sampel pada data uji. Kondisi ini berpotensi menyebabkan model mengalami kesulitan dalam mempelajari karakteristik kelas minoritas sehingga memengaruhi performa klasifikasi pada kategori tersebut. Oleh karena itu, penelitian ini menggunakan Focal Loss sebagai salah satu skenario pengujian untuk menganalisis pengaruhnya terhadap kemampuan model dalam menangani ketidakseimbangan kelas.

### 4.1.2 Pemrosesan Data

Tahapan *preprocessing* data dalam penelitian ini dilakukan melalui beberapa langkah utama untuk menghasilkan data yang bersih dan terstruktur sebelum digunakan dalam proses pelatihan model. Tahapan tersebut meliputi case folding, yaitu mengubah seluruh huruf dalam teks menjadi huruf kecil (lowercase), serta pembersihan karakter non-alfabet seperti tanda baca dan simbol yang tidak relevan. Proses ini bertujuan untuk mengurangi noise pada data dan meningkatkan kualitas representasi teks.

Selanjutnya, dilakukan proses tokenisasi menggunakan tokenizer IndoBERT yang berfungsi untuk mengubah teks menjadi token-token yang sesuai dengan vocabulary model. Setelah itu, dilakukan proses padding dan truncation untuk menyeragamkan panjang input sesuai dengan max sequence length yang telah ditentukan, yaitu sepanjang 512 token. Tahap terakhir adalah encoding label ke dalam bentuk numerik agar dapat diproses oleh model dalam proses klasifikasi.

Setelah seluruh tahapan preprocessing selesai, data dikonversi ke dalam bentuk tensor dan digunakan sebagai input pada model IndoBERT. Proses fine-tuning dilakukan dengan menggunakan optimizer AdamW. Proses ini bertujuan untuk menyesuaikan parameter model agar dapat mengenali pola spesifik pada dataset yang digunakan, sehingga mampu menghasilkan performa klasifikasi yang optimal.

Contoh hasil tokenisasi data ditampilkan pada Tabel 4.2, yang menunjukkan perubahan teks dari bentuk mentah hingga menjadi teks yang siap digunakan.

**Tabel 4. 2 Tokenisasi data**

| Input | Output Token |
|---|---|
| Video Jet Tempur Rusia tiba di Biak Papua | ['video', 'jet', 'tempur', 'rusia', 'tiba', 'di', 'biak', 'papua'] |

Tabel tokenisasi menunjukkan proses pemecahan teks menjadi unit-unit kecil (token) menggunakan tokenizer dari *IndoBERT*, di mana teks input yang telah melalui tahap preprocessing diubah menjadi token berbasis metode *subword (WordPiece)*. Hasil tokenisasi memperlihatkan bahwa kata-kata umum seperti "video", "jet", "tempur", dan "rusia" dapat dikenali dengan baik, sementara beberapa kata dipecah menjadi subword seperti "pengebom" menjadi "pengeb" dan "##om". Selain itu, munculnya token [UNK] menunjukkan adanya kata yang tidak terdapat dalam kosakata model, yang umumnya disebabkan oleh nama entitas, kata yang jarang digunakan, atau kompleksitas teks yang terlalu panjang.

## 4.2 Hasil Pengujian

Pada bagian ini disajikan hasil pengujian model klasifikasi teks menggunakan *IndoBERT* dalam mendeteksi konten hoaks dengan dua kategori, yaitu SALAH dan PENIPUAN. Pengujian dilakukan berdasarkan skenario yang telah dijelaskan pada Subbab 4.1, yaitu dengan membandingkan dua pendekatan fungsi loss, yaitu Cross Entropy sebagai baseline dan Focal Loss untuk mengatasi ketidakseimbangan data.

Evaluasi performa model dilakukan menggunakan beberapa metrik, yaitu akurasi (*accuracy*), presisi (*precision*), *recall*, dan F1-score. Selain itu, digunakan pula *confusion matrix* untuk melihat distribusi kesalahan klasifikasi pada masing-

masing kelas. Hasil pengujian ditampilkan dalam bentuk tabel dan confusion matrix untuk memudahkan analisis dan perbandingan performa kedua pendekatan.

### 4.2.1 Hasil Pengujian Skenario Klasifikasi Biner

Pada skenario pertama, pengujian dilakukan menggunakan pendekatan klasifikasi biner dengan dua kategori, yaitu SALAH dan PENIPUAN. Dataset yang digunakan terdiri dari 970 data hasil filtering, yang meliputi 698 data SALAH dan 272 data PENIPUAN. Seluruh data telah melalui tahapan preprocessing yang meliputi pembersihan teks, normalisasi, tokenisasi menggunakan tokenizer IndoBERT, serta encoding label ke dalam bentuk numerik.

#### 4.2.1.1 Pengujian Menggunakan *Cross Entropy (Baseline*)

Percobaan pertama dilakukan dengan menggunakan fungsi *loss Cross Entropy* sebagai model *baseline*. Model dilatih menggunakan data *training* dan diuji menggunakan data *testing* dengan rasio pembagian data sebesar 70:15:15, di mana 70 persen data digunakan untuk pelatihan, 15 persen untuk validasi, dan 15 persen untuk pengujian. Dataset yang digunakan merupakan dataset klasifikasi biner dengan dua label, yaitu SALAH dan PENIPUAN. Seluruh data telah melalui tahap *preprocessing* dan tokenisasi menggunakan IndoBERT yang menghasilkan representasi teks dalam bentuk token yang siap digunakan oleh model.

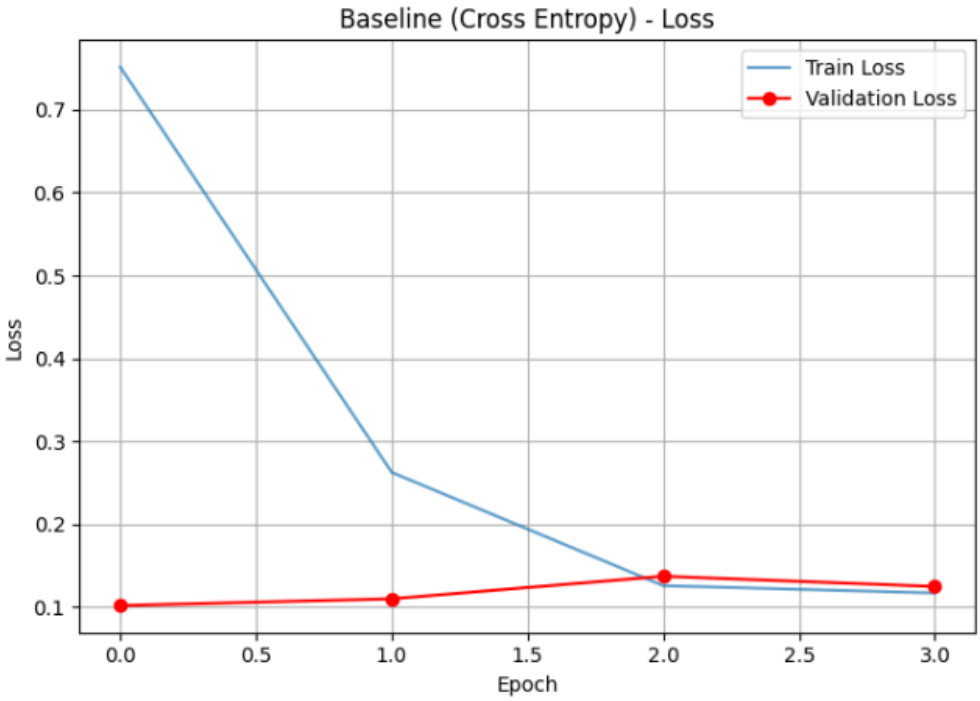
Proses pelatihan model dilakukan hingga 10 *epoch* dengan mekanisme *early stopping*, sehingga pelatihan berhenti lebih awal ketika performa model tidak mengalami peningkatan signifikan. *Batch size* yang digunakan adalah sebesar 16.

Selama pelatihan, model melakukan penyesuaian bobot berdasarkan nilai *loss* yang dihasilkan. Berdasarkan hasil pelatihan, model *baseline* mencapai F1-score macro validasi terbaik sebesar 97,39% pada *epoch* ke-1. Evaluasi model dilakukan menggunakan metrik akurasi, *precision*, *recall* dan F1-score, serta *confusion matrix* untuk melihat distribusi hasil klasifikasi pada masing-masing kelas.

**Tabel 4. 3 Hasil Pengujian Baseline Biner**

| **Metrik** | **Nilai** |
|---|---|
| Akurasi | 97,26% |
| F1-Score (Macro) | 96,56% |
| F1-Score (SALAH) | 98,11% |
| F1-Score (PENIPUAN) | 95,00% |
| Loss | 0,0923 |

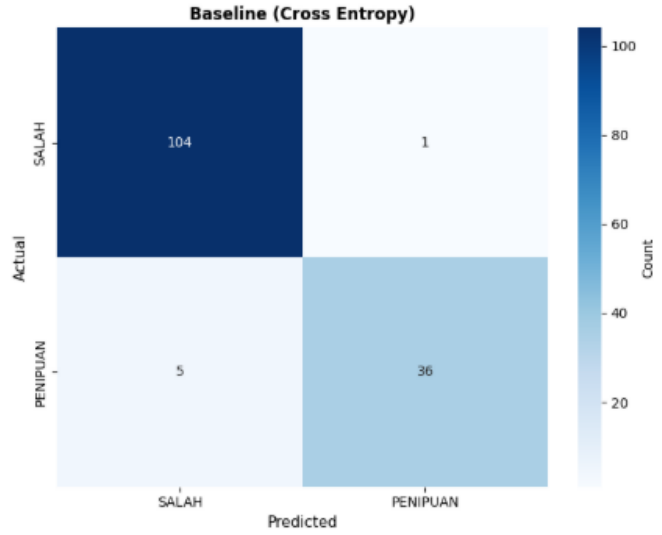
Berdasarkan Tabel 4.3, model *baseline* menunjukkan performa yang baik dengan nilai akurasi sebesar 97,26% dan F1-score macro sebesar 96,56%, yang mengindikasikan bahwa secara keseluruhan model mampu mengklasifikasikan kedua kelas dengan tingkat keseimbangan yang baik. Namun, jika dianalisis lebih lanjut pada masing-masing kelas, terlihat adanya perbedaan performa antara kelas SALAH dan PENIPUAN. Nilai F1-score pada kelas SALAH sebesar 98,11% lebih tinggi dibandingkan kelas PENIPUAN yang mencapai 95,00%, dengan selisih sekitar 3,11%. Perbedaan ini menunjukkan bahwa model lebih optimal dalam mengenali pola pada kelas SALAH, yang dapat disebabkan oleh distribusi data yang tidak seimbang atau kompleksitas karakteristik teks pada kelas PENIPUAN yang lebih sulit dipelajari oleh model.




**Gambar 4. 1 Grafik Loss Model Baseline Biner**

Pada Gambar 4.1, grafik loss model baseline menunjukkan bahwa nilai training loss mengalami penurunan yang cukup signifikan dari epoch awal hingga akhir, yang menandakan bahwa model mampu mempelajari pola data latih dengan baik. Namun, validation loss cenderung tidak mengalami penurunan yang konsisten dan bahkan sedikit meningkat pada epoch-epoch tertentu, yang mengindikasikan adanya potensi overfitting. Kondisi ini menunjukkan bahwa model mulai kehilangan kemampuan generalisasi terhadap data di luar data latih. Hal tersebut juga diperkuat oleh grafik F1-score yang pada awalnya stabil tetapi mengalami penurunan pada epoch terakhir. Penurunan nilai F1-score ini mengindikasikan bahwa performa model terhadap data validasi tidak terjaga secara optimal. Dengan demikian, meskipun model baseline menunjukkan

pembelajaran yang baik pada data latih, kestabilan performanya pada data validasi masih perlu ditingkatkan.




**Gambar 4. 2 Confusion Matrix *Baseline* Biner**

Berdasarkan confusion matrix pada Gambar 4.2, diperoleh nilai True Negative (SALAH diprediksi sebagai SALAH) sebanyak 104 data, False Positive (PENIPUAN diprediksi sebagai SALAH) sebanyak 1 data, False Negative (SALAH diprediksi sebagai PENIPUAN) sebanyak 3 data, dan True Positive (PENIPUAN diprediksi sebagai PENIPUAN) sebanyak 38 data. Secara umum, sebagian besar prediksi model berada pada diagonal utama, yang mengindikasikan bahwa mayoritas data berhasil diklasifikasikan dengan benar. Hal ini menunjukkan bahwa model memiliki kemampuan yang cukup baik dalam membedakan kedua kelas.

Namun demikian, masih terdapat kesalahan klasifikasi yang perlu diperhatikan, khususnya pada nilai false negative yang mencapai 3 data, di mana

data dengan label SALAH diprediksi sebagai PENIPUAN. Kondisi ini menunjukkan bahwa model masih mengalami kesulitan dalam mengenali karakteristik tertentu dari kelas SALAH sehingga cenderung keliru mengklasifikasikannya. Di sisi lain, nilai false positive yang relatif kecil, yaitu hanya 1 data, mengindikasikan bahwa model jarang salah dalam mengklasifikasikan data PENIPUAN sebagai SALAH. Dengan demikian, dapat disimpulkan bahwa model memiliki tingkat presisi yang cukup tinggi dalam mendeteksi kelas PENIPUAN, meskipun kemampuan recall-nya belum sepenuhnya optimal.

#### 4.2.1.2 Pengujian Menggunakan *Focal Loss*

Percobaan kedua dilakukan dengan menggunakan fungsi loss Focal Loss untuk meningkatkan performa model dalam menangani data yang sulit diklasifikasikan. Model dilatih dan diuji menggunakan rasio pembagian data yang sama, yaitu 70:15:15, dengan dataset yang identik seperti pada pengujian baseline. Parameter Focal Loss yang digunakan adalah alpha sebesar 0,25 untuk kelas SALAH dan 0,75 untuk kelas PENIPUAN serta gamma sebesar 2,0. Parameter alpha berfungsi sebagai penyeimbang kelas di mana kelas PENIPUAN sebagai kelas minoritas diberikan bobot lebih besar, sedangkan parameter gamma mengontrol seberapa besar pengurangan loss pada sampel yang mudah diklasifikasikan. Proses pelatihan dilakukan hingga maksimal 10 epoch dengan batch size sebesar 16 serta menggunakan mekanisme *early stopping*. Focal Loss bekerja dengan memberikan bobot lebih besar pada data yang sulit

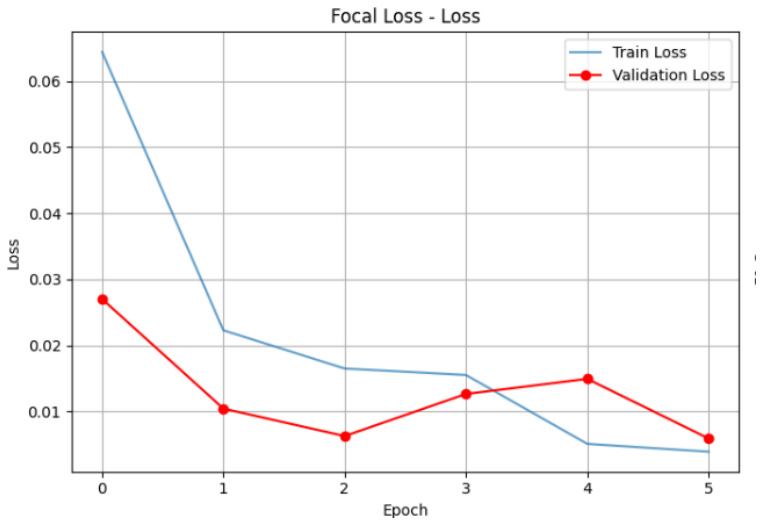
diklasifikasikan sehingga model dapat lebih fokus pada kesalahan selama proses pelatihan.

Berdasarkan hasil pelatihan, model Focal Loss mencapai F1-score macro validasi terbaik sebesar 98,27% pada epoch ke-3.

**Tabel 4. 4 Hasil Pengujian Focal Loss Biner**

| **Metrik** | **Nilai** |
|---|---|
| Accuracy | 98,63% |
| F1-Score (Macro) | 98,33% |
| F1-Score (SALAH) | 99,04% |
| F1-Score (PENIPUAN) | 97,62% |
| Loss | 0,0059 |

Berdasarkan Tabel 4.4, model dengan *Focal Loss* menunjukkan peningkatan performa yang signifikan dibandingkan model *baseline*. Nilai akurasi meningkat menjadi 98,63% (naik 1,37%), *F1-score macro* meningkat menjadi 98,33% (naik 1,77%), F1-score untuk kelas SALAH mencapai 99,04% (naik 0,93%), dan F1-score untuk kelas PENIPUAN mencapai 97,62% (naik **2,62%**). Selain itu, nilai *loss* berhasil ditekan secara drastis hingga mencapai 0,0059, atau turun sebesar 93,61% dibandingkan model *baseline* (dari 0,0923 menjadi 0,0059). Peningkatan ini menunjukkan bahwa *Focal Loss* efektif dalam membantu model mempelajari distribusi data yang lebih kompleks, terutama pada kelas minoritas.




**Gambar 4. 3 Grafik loss model *Focal Loss* Biner**

Pada Gambar 4.3, grafik loss model dengan Focal Loss menunjukkan bahwa baik *training loss* maupun *validation loss* mengalami penurunan yang konsisten. Training loss turun drastis dari 0,0223 pada *epoch* pertama menjadi 0,0001 pada *epoch* kelima, sementara *validation loss* berada pada rentang yang relatif stabil (0,0270 pada *epoch* pertama, 0,0104 pada *epoch* kedua, 0,0063 pada *epoch* ketiga, kemudian sedikit meningkat menjadi 0,0127 dan 0,0149 pada *epoch* keempat dan kelima). Hal ini menunjukkan bahwa model tidak hanya mampu mempelajari data latih dengan baik, tetapi juga memiliki kemampuan generalisasi yang lebih stabil terhadap data validasi dibandingkan model baseline yang menunjukkan validation loss meningkat hingga 0,1233. Dibandingkan dengan model baseline, Focal Loss mampu menjaga keseimbangan antara proses pembelajaran dan generalisasi model, sehingga terbukti memberikan performa yang lebih stabil dan optimal dalam proses klasifikasi.


**Gambar 4. 4 Confusion matrix *Focal Loss* Biner**

Berdasarkan confusion matrix pada Gambar 4.4, diperoleh nilai True Negative (SALAH diprediksi sebagai SALAH) sebanyak 103 data, False Positive (PENIPUAN diprediksi sebagai SALAH) sebanyak 2 data, False Negative (SALAH diprediksi sebagai PENIPUAN) sebesar 0 data, dan True Positive (PENIPUAN diprediksi sebagai PENIPUAN) sebanyak 41 data. Distribusi ini menunjukkan bahwa sebagian besar prediksi model berada pada diagonal utama, yang menandakan tingkat klasifikasi yang sangat baik.

Keberhasilan model dalam menghilangkan seluruh kesalahan false negative (dari 3 data pada model baseline menjadi 0 data) menjadi indikasi bahwa model mampu mengenali seluruh data dari kelas SALAH dengan tepat. Hal ini mencerminkan peningkatan kemampuan model dalam membedakan karakteristik antar kelas dibandingkan pendekatan sebelumnya. Dengan tidak adanya false negative, nilai recall untuk kelas SALAH mencapai tingkat optimal (100%), yang

mengindikasikan bahwa model memiliki sensitivitas yang sangat baik dalam mengenali seluruh data pada kelas tersebut. Peningkatan ini menunjukkan efektivitas penggunaan Focal Loss dalam menangani distribusi data yang sebelumnya sulit dipelajari oleh model.

Meskipun demikian, terdapat sedikit peningkatan pada nilai false positive, yaitu sebanyak 2 data (dari 1 data menjadi 2 data), di mana data PENIPUAN diprediksi sebagai SALAH. Namun, peningkatan ini tergolong tidak signifikan jika dibandingkan dengan peningkatan performa yang dihasilkan, khususnya dalam menghilangkan false negative. Kesalahan ini menunjukkan bahwa model masih memiliki keterbatasan dalam membedakan beberapa karakteristik tertentu dari kelas PENIPUAN, namun jumlah kesalahan yang relatif kecil menunjukkan bahwa presisi model tetap berada pada tingkat yang baik. Dengan demikian, peningkatan false positive masih dapat ditoleransi karena memberikan dampak positif yang lebih besar terhadap performa keseluruhan model, terutama dalam konteks deteksi hoaks di mana false negative (berita SALAH yang salah diprediksi sebagai PENIPUAN) memiliki konsekuensi yang lebih berbahaya.

### 4.2.2 Hasil Pengujian Skenario Klasifikasi Multikelas

Pada skenario kedua, pengujian dilakukan menggunakan pendekatan klasifikasi multikelas dengan tiga kategori, yaitu SALAH, PENIPUAN, dan OTHERS. Kategori OTHERS merupakan hasil penggabungan beberapa kategori yang memiliki jumlah data relatif sedikit, yaitu BELUM TERBUKTI, SATIR,

dan PARODI. Setelah proses penggabungan dilakukan, diperoleh total 987 data yang terdiri dari 698 data SALAH, 272 data PENIPUAN, dan 17 data OTHERS.

Seluruh data telah melalui tahapan preprocessing yang sama dengan skenario klasifikasi biner, yaitu pembersihan teks, normalisasi, tokenisasi menggunakan tokenizer IndoBERT, serta encoding label ke dalam bentuk numerik. Pembagian dataset dilakukan menggunakan metode stratified split dengan rasio 70:15:15 untuk data training, validation, dan testing sehingga distribusi setiap kelas tetap proporsional pada seluruh subset data.

Penambahan kategori OTHERS bertujuan untuk menganalisis pengaruh jumlah kelas terhadap performa model serta menguji kemampuan IndoBERT dalam mengklasifikasikan data dengan distribusi yang lebih tidak seimbang. Dibandingkan skenario klasifikasi biner, skenario ini memiliki tingkat ketidakseimbangan kelas yang lebih tinggi karena jumlah data pada kategori OTHERS jauh lebih sedikit dibandingkan kategori SALAH dan PENIPUAN.

Seperti pada skenario sebelumnya, pengujian dilakukan menggunakan dua fungsi loss, yaitu Cross Entropy sebagai model baseline dan Focal Loss sebagai pendekatan untuk menangani ketidakseimbangan kelas. Evaluasi performa model dilakukan menggunakan metrik Accuracy, Precision, Recall, dan F1-Score, serta Confusion Matrix untuk menganalisis distribusi prediksi pada masing-masing kelas. Hasil pengujian untuk kedua fungsi loss tersebut dijelaskan pada subbab berikut.

#### 4.2.2.1 Pengujian Menggunakan *Cross Entropy (Baseline*)

Percobaan pertama pada skenario klasifikasi multikelas dilakukan menggunakan fungsi loss Cross Entropy sebagai model baseline. Model dilatih menggunakan data training dan dievaluasi menggunakan data validation serta data testing dengan rasio pembagian data sebesar 70:15:15. Dataset yang digunakan terdiri dari tiga label, yaitu SALAH, PENIPUAN, dan OTHERS. Seluruh data telah melalui tahap preprocessing dan tokenisasi menggunakan tokenizer IndoBERT sehingga dapat diproses oleh model dalam bentuk representasi numerik.

Proses pelatihan dilakukan dengan maksimum 10 epoch menggunakan mekanisme early stopping untuk mencegah terjadinya overfitting. Batch size yang digunakan adalah 16. Selama proses pelatihan, model melakukan pembaruan parameter berdasarkan nilai loss yang dihasilkan oleh fungsi Cross Entropy. Berdasarkan hasil pelatihan, model memperoleh nilai F1-score macro validasi terbaik sebesar 64,41% pada epoch ke-3. Selanjutnya, model dievaluasi menggunakan data testing untuk mengetahui performa akhir klasifikasi.

**Tabel 4. 5 Hasil Pengujian Baseline pada skenario multikelas**

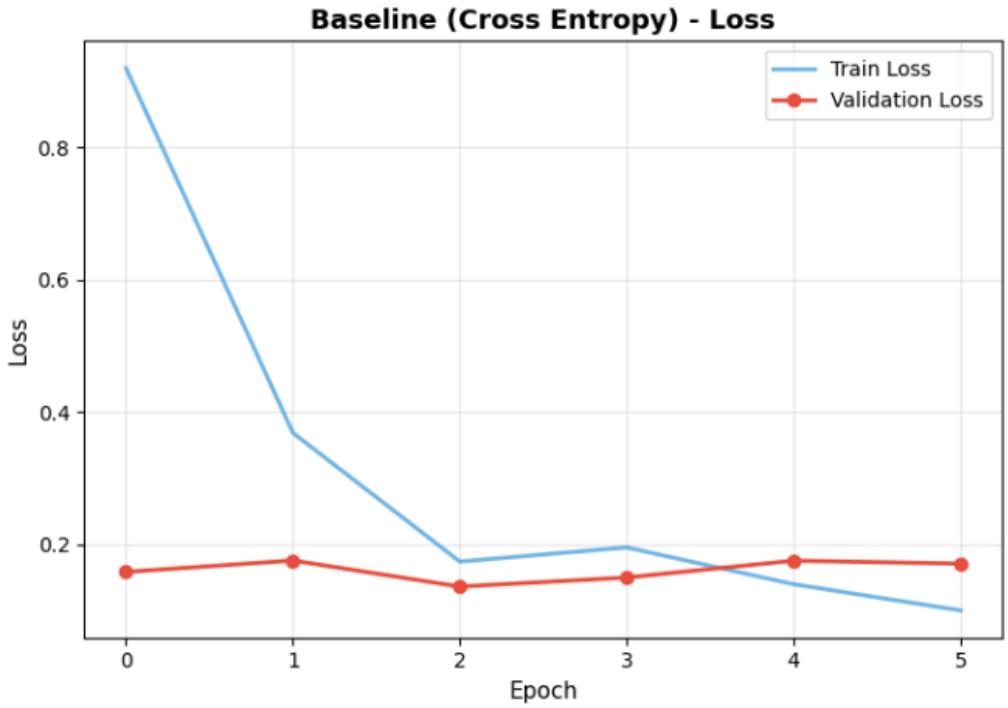
| Metrik | Nilai |
|---|---|
| Accuracy | 96,64% |
| F1-Score (Macro) | 65,06% |
| F1-Score (SALAH) | 97,67% |
| F1-Score (PENIPUAN) | 97,50% |
| F1-Score (OTHERS) | 0,00% |
| Loss | 0,1759 |

Berdasarkan Tabel 4.5, model baseline menghasilkan akurasi sebesar 96,64% dengan F1-score macro sebesar 65,06%. Nilai akurasi yang tinggi

menunjukkan bahwa sebagian besar data berhasil diklasifikasikan dengan benar. Namun, nilai F1-score macro yang jauh lebih rendah dibandingkan akurasi mengindikasikan adanya ketidakseimbangan performa antar kelas.

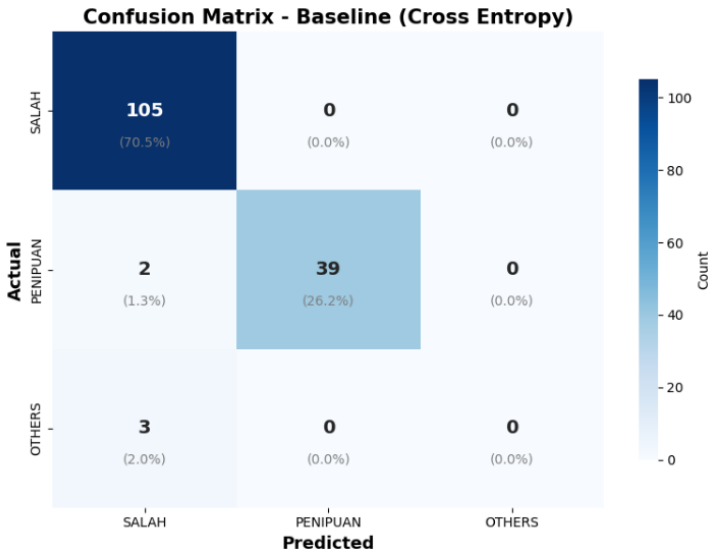
Jika dianalisis berdasarkan masing-masing kategori, model menunjukkan performa yang sangat baik pada kelas SALAH dan PENIPUAN dengan F1-score masing-masing sebesar 97,67% dan 97,50%. Sebaliknya, model gagal mengklasifikasikan kelas OTHERS yang ditunjukkan oleh nilai F1-score sebesar 0,00%. Hasil ini menunjukkan bahwa model belum mampu mempelajari karakteristik kelas OTHERS secara memadai.

Rendahnya performa pada kelas OTHERS dapat disebabkan oleh jumlah data yang sangat sedikit dibandingkan kelas lainnya. Pada dataset yang digunakan, kelas OTHERS hanya memiliki 17 sampel dari total 987 data, sehingga kontribusinya terhadap proses pelatihan menjadi sangat kecil. Akibatnya, model cenderung memprioritaskan pembelajaran pada kelas mayoritas, yaitu SALAH dan PENIPUAN, yang memiliki jumlah data jauh lebih besar.




**Gambar 4. 5 Grafik Loss Model *Baseline* Multikelas**

Pada Gambar 4.5, terlihat bahwa training loss mengalami penurunan yang cukup signifikan selama proses pelatihan, yang menunjukkan bahwa model mampu mempelajari pola pada data training dengan baik. Validation loss juga cenderung berada pada rentang yang relatif stabil meskipun mengalami sedikit fluktuasi pada beberapa epoch. Hal ini menunjukkan bahwa model memiliki kemampuan generalisasi yang cukup baik terhadap data validasi.




**Gambar 4. 6 Confusion Matrix *Baseline* Multikelas**

Berdasarkan confusion matrix pada Gambar 4.6, diperoleh bahwa seluruh data pada kelas SALAH berhasil diklasifikasikan dengan benar, yaitu sebanyak 105 data. Pada kelas PENIPUAN, sebanyak 39 data berhasil diklasifikasikan dengan benar, sedangkan 2 data salah diklasifikasikan sebagai SALAH. Sementara itu, seluruh data pada kelas OTHERS sebanyak 3 data salah diklasifikasikan sebagai SALAH dan tidak ada satupun yang berhasil dikenali sebagai kelas OTHERS.

Hasil ini menunjukkan bahwa model memiliki kemampuan yang sangat baik dalam membedakan kelas SALAH dan PENIPUAN, tetapi belum mampu mengenali karakteristik kelas OTHERS. Ketidakseimbangan distribusi data yang sangat tinggi menyebabkan model lebih fokus mempelajari pola pada kelas mayoritas sehingga kelas minoritas tidak memperoleh representasi yang cukup selama proses pelatihan. Akibatnya, meskipun nilai akurasi tetap tinggi, performa

model pada klasifikasi multikelas belum dapat dikatakan optimal karena masih gagal mengidentifikasi kelas OTHERS.

#### 4.2.2.2 Pengujian Menggunakan *Focal Loss*

Percobaan kedua pada skenario klasifikasi tiga kelas dilakukan dengan menggunakan fungsi loss Focal Loss. Penggunaan Focal Loss bertujuan untuk meningkatkan kemampuan model dalam menangani ketidakseimbangan distribusi data, khususnya karena kelas OTHERS hanya memiliki jumlah sampel yang sangat sedikit dibandingkan kelas SALAH dan PENIPUAN. Pada penelitian ini, parameter Focal Loss menggunakan nilai gamma sebesar 2,0 dan bobot kelas (alpha) yang disesuaikan berdasarkan distribusi data masing-masing kelas.

Proses pelatihan dilakukan menggunakan konfigurasi yang sama seperti pada model baseline, yaitu maksimum 10 epoch, batch size sebesar 16, optimizer AdamW, serta mekanisme early stopping untuk mencegah overfitting. Berdasarkan hasil pelatihan, model mencapai nilai F1-score macro validasi terbaik sebesar 64,99% pada epoch ke-3.

Hasil pengujian model Focal Loss pada data testing ditunjukkan pada Tabel 4.4 dibawah.

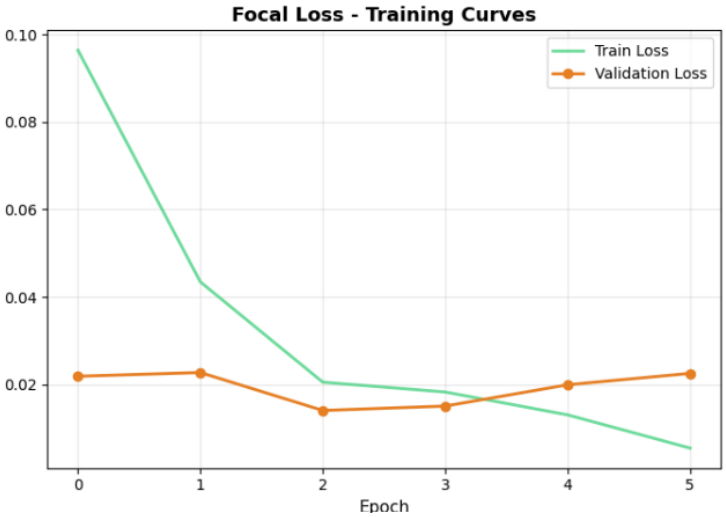
**Tabel 4. 5 Hasil Pengujian Focal Loss Multikelas**

| Metrik | Nilai |
|---|---|
| Accuracy | 95,97% |
| F1-Score (Macro) | 64,47% |
| F1-Score (SALAH) | 97,22% |
| F1-Score (PENIPUAN) | 96,20% |
| F1-Score (OTHERS) | 0,00% |
| Loss | 0,0225 |

Berdasarkan Tabel 4.4, model model Focal Loss memperoleh nilai akurasi sebesar 95,97% dengan F1-score macro sebesar 64,47%. Nilai F1-score untuk kelas SALAH mencapai 97,22%, sedangkan kelas PENIPUAN memperoleh F1-score sebesar 96,20%. Namun, seperti halnya pada model baseline, kelas OTHERS memperoleh nilai F1-score sebesar 0,00%.

Meskipun Focal Loss dirancang untuk memberikan perhatian lebih besar pada sampel yang sulit dipelajari dan kelas minoritas, jumlah data OTHERS yang sangat sedikit menyebabkan model belum mampu mempelajari karakteristik kelas tersebut secara optimal. Akibatnya, seluruh data OTHERS pada data uji gagal dikenali oleh model dan diklasifikasikan sebagai kelas SALAH.

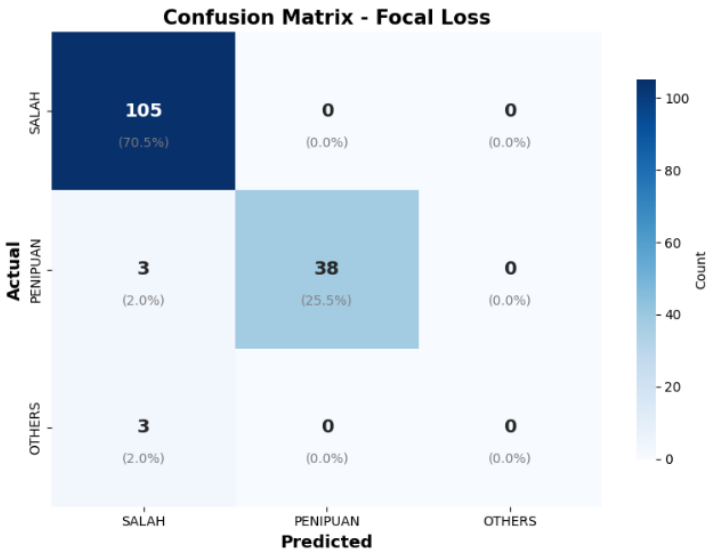
Jika dibandingkan dengan model baseline, penggunaan Focal Loss menghasilkan nilai loss yang jauh lebih rendah, yaitu 0,0225 dibandingkan 0,1706 pada Cross Entropy. Penurunan loss ini menunjukkan bahwa proses optimasi model berjalan lebih baik selama pelatihan. Akan tetapi, penurunan nilai loss tersebut tidak diikuti dengan peningkatan performa klasifikasi secara keseluruhan. Nilai akurasi menurun dari 96,64% menjadi 95,97%, sedangkan F1-score macro turun dari 65,06% menjadi 64,47%.




**Gambar 4. 7 Grafik loss model *Focal Loss* Multikelas**

Pada Gambar 4.7, grafik loss menggambarkan bahwa training loss mengalami penurunan yang sangat signifikan dari 0,0434 pada epoch pertama menjadi 0,0004 pada epoch kelima. Validation loss juga berada pada nilai yang relatif rendah dengan nilai terbaik sebesar 0,0140 pada epoch ketiga. Hasil ini menunjukkan bahwa model mampu mempelajari pola data latih dengan baik dan menghasilkan proses optimasi yang stabil selama pelatihan.

Namun demikian, nilai F1-score macro tidak menunjukkan peningkatan yang signifikan. Hal ini mengindikasikan bahwa rendahnya performa model bukan disebabkan oleh kegagalan proses pelatihan, melainkan karena keterbatasan jumlah sampel pada kelas OTHERS yang terlalu sedikit sehingga model tidak memiliki cukup informasi untuk mempelajari pola kelas tersebut.




**Gambar 4. 8 Confusion matrix *Focal Loss* Multikelas**

Berdasarkan confusion matrix pada Gambar 4.8, diperoleh 105 data kelas SALAH berhasil diprediksi dengan benar sebagai SALAH. Pada kelas PENIPUAN, sebanyak 38 data berhasil diprediksi dengan benar, sedangkan 3 data salah diklasifikasikan sebagai SALAH. Untuk kelas OTHERS, seluruh 3 data gagal dikenali oleh model dan diprediksi sebagai SALAH.

Hasil confusion matrix menunjukkan bahwa model memiliki kemampuan yang sangat baik dalam membedakan kelas SALAH dan PENIPUAN. Namun, model belum mampu mengenali kelas OTHERS sama sekali, yang ditunjukkan oleh nilai recall dan precision sebesar 0%. Kondisi ini menyebabkan nilai F1-score kelas OTHERS juga menjadi 0%, sehingga menurunkan nilai F1-score macro secara keseluruhan.

Secara umum, hasil pengujian menunjukkan bahwa penggunaan Focal Loss pada skenario klasifikasi tiga kelas belum mampu memberikan peningkatan performa dibandingkan Cross Entropy. Meskipun berhasil menghasilkan nilai loss yang lebih rendah, model tetap mengalami kesulitan dalam mempelajari kelas OTHERS yang memiliki jumlah sampel sangat terbatas. Oleh karena itu, pada kondisi distribusi data seperti penelitian ini, jumlah data pada kelas minoritas menjadi faktor yang lebih berpengaruh terhadap performa model dibandingkan pemilihan fungsi loss yang digunakan.

### 4.2.3 Perbandingan Hasil Pengujian Antar Skenario

Untuk mengetahui pengaruh penambahan kelas OTHERS terhadap performa model, dilakukan perbandingan antara skenario klasifikasi biner dan klasifikasi tiga kelas. Selain itu, dilakukan pula analisis terhadap penggunaan fungsi loss Cross Entropy dan Focal Loss pada masing-masing skenario. Perbandingan ini bertujuan untuk mengevaluasi perubahan performa model ketika kompleksitas tugas klasifikasi ditingkatkan dari dua kelas menjadi tiga kelas.

Tabel 4.6 menyajikan ringkasan hasil pengujian dari seluruh skenario yang digunakan dalam penelitian.

**Tabel 4. 6 Ringkasan hasil pengujian dua skenario**

| Skenario | Accuracy | F1 Macro | F1 SALAH | F1 PENIPUAN | F1 OTHERS |
|---|---|---|---|---|---|
| Biner - Cross Entropy | 97,26% | 96,56% | 98,11% | 95,00% | - |
| Biner - Focal Loss | 98,63% | 98,33% | 99,04% | 97,62% | - |
| 3 Kelas - Cross Entropy | 96,64% | 65,06% | 97,67% | 97,50% | 0,00% |
| 3 Kelas - Focal Loss | 95,97% | 64,47% | 97,22% | 96,20% | 0,00% |

Berdasarkan Tabel 4.6, terlihat bahwa performa terbaik diperoleh pada skenario klasifikasi biner menggunakan Focal Loss dengan nilai akurasi sebesar 98,63% dan F1-score macro sebesar 98,33%. Hasil ini menunjukkan bahwa Focal Loss efektif dalam meningkatkan kemampuan model untuk mengenali kelas minoritas PENIPUAN pada skenario dua kelas.

Ketika kelas OTHERS ditambahkan ke dalam dataset, terjadi penurunan performa pada kedua fungsi loss. Nilai F1-score macro turun secara signifikan dari 96,56% menjadi 65,06% pada Cross Entropy dan dari 98,33% menjadi 64,47% pada Focal Loss. Penurunan ini terutama disebabkan oleh kegagalan model dalam mengenali kelas OTHERS, yang ditunjukkan oleh nilai F1-score sebesar 0,00% pada kedua pendekatan.

Meskipun demikian, performa pada kelas SALAH dan PENIPUAN tetap berada pada tingkat yang sangat tinggi. Pada skenario tiga kelas menggunakan Cross Entropy, F1-score kelas SALAH mencapai 97,67% dan kelas PENIPUAN mencapai 97,50%. Sementara itu, pada Focal Loss diperoleh F1-score sebesar 97,22% untuk kelas SALAH dan 96,20% untuk kelas PENIPUAN. Hasil ini menunjukkan bahwa penambahan kelas OTHERS tidak memberikan dampak yang signifikan terhadap kemampuan model dalam membedakan kelas SALAH dan PENIPUAN.

Penurunan nilai F1-score macro terjadi karena metrik tersebut menghitung rata-rata performa seluruh kelas secara seimbang. Dengan nilai F1-score kelas OTHERS sebesar 0,00%, rata-rata keseluruhan menjadi turun meskipun performa pada dua kelas utama tetap tinggi. Kondisi ini mengindikasikan bahwa jumlah

data OTHERS yang sangat sedikit belum cukup untuk membentuk representasi pola yang dapat dipelajari secara efektif oleh model.

Selain itu, hasil penelitian menunjukkan bahwa penggunaan Focal Loss tidak selalu menghasilkan performa yang lebih baik dibandingkan Cross Entropy. Pada skenario klasifikasi biner, Focal Loss berhasil meningkatkan performa model. Namun pada skenario tiga kelas, Cross Entropy justru memperoleh akurasi dan F1-score macro yang sedikit lebih tinggi. Temuan ini menunjukkan bahwa efektivitas Focal Loss sangat dipengaruhi oleh distribusi data dan jumlah sampel pada setiap kelas.

Secara keseluruhan, hasil pengujian membuktikan bahwa model IndoBERT memiliki kemampuan yang sangat baik dalam mengklasifikasikan berita ke dalam kategori SALAH dan PENIPUAN. Akan tetapi, penambahan kelas OTHERS dengan jumlah data yang sangat terbatas menyebabkan model mengalami kesulitan dalam mempelajari karakteristik kelas tersebut. Oleh karena itu, peningkatan jumlah data pada kelas OTHERS menjadi faktor yang lebih penting dibandingkan perubahan fungsi loss untuk meningkatkan performa klasifikasi tiga kelas.

### 4.3 Pembahasan

Berdasarkan hasil pengujian yang telah dilakukan, model IndoBERT menunjukkan performa yang baik dalam melakukan klasifikasi berita hoaks pada seluruh skenario yang digunakan dalam penelitian ini. Pengujian dilakukan pada dua skenario utama, yaitu klasifikasi biner dan klasifikasi tiga kelas, dengan menggunakan fungsi loss Cross Entropy dan Focal Loss. Evaluasi model

dilakukan menggunakan metrik accuracy, precision, recall, F1-score, serta confusion matrix. Hasil pengujian menunjukkan adanya perbedaan performa antara kedua fungsi loss pada masing-masing skenario klasifikasi.

Pada skenario klasifikasi biner, penggunaan Focal Loss menghasilkan performa yang lebih baik dibandingkan Cross Entropy. Model dengan Cross Entropy memperoleh akurasi sebesar 97,26% dan F1-score macro sebesar 96,56%, sedangkan model dengan Focal Loss memperoleh akurasi sebesar 98,63% dan F1-score macro sebesar 98,33%. Peningkatan performa juga terlihat pada masing-masing kelas, di mana F1-score kelas SALAH meningkat dari 98,11% menjadi 99,04% dan F1-score kelas PENIPUAN meningkat dari 95,00% menjadi 97,62%. Selain itu, nilai loss pada model Focal Loss lebih rendah dibandingkan model Cross Entropy, yang menunjukkan proses pembelajaran model berlangsung dengan lebih baik.

Hasil confusion matrix pada skenario klasifikasi biner memperlihatkan bahwa sebagian besar data berhasil diklasifikasikan dengan benar oleh kedua model. Pada model baseline, masih terdapat beberapa data yang salah diklasifikasikan sehingga menghasilkan nilai false positive dan false negative. Setelah menggunakan Focal Loss, jumlah kesalahan klasifikasi berkurang dan distribusi prediksi benar pada diagonal utama confusion matrix menjadi lebih dominan. Kondisi ini menunjukkan bahwa model mampu membedakan karakteristik antara kelas SALAH dan PENIPUAN dengan tingkat ketepatan yang lebih tinggi. Hasil tersebut juga sejalan dengan peningkatan nilai precision, recall, dan F1-score yang diperoleh pada kedua kelas.

Pada skenario klasifikasi tiga kelas, hasil pengujian menunjukkan pola yang berbeda dibandingkan skenario klasifikasi biner. Model Cross Entropy memperoleh nilai akurasi sebesar 96,64% dengan F1-score macro sebesar 65,06%, sedangkan model Focal Loss memperoleh akurasi sebesar 95,97% dengan F1-score macro sebesar 64,47%. Berdasarkan hasil tersebut, model Cross Entropy menghasilkan performa yang sedikit lebih tinggi dibandingkan model Focal Loss. Meskipun demikian, kedua model tetap menunjukkan performa yang sangat baik pada kelas SALAH dan PENIPUAN dengan nilai F1-score di atas 96%.

Berdasarkan confusion matrix pada skenario tiga kelas, seluruh data pada kelas OTHERS tidak berhasil diprediksi dengan benar oleh kedua model. Pada model Cross Entropy, ketiga data OTHERS pada data uji diprediksi sebagai kelas SALAH. Hasil yang sama juga diperoleh pada model Focal Loss, di mana seluruh data OTHERS kembali diprediksi sebagai SALAH. Akibatnya, nilai precision, recall, dan F1-score untuk kelas OTHERS bernilai 0,00% pada kedua model. Kondisi tersebut menunjukkan bahwa tidak terdapat prediksi yang benar untuk kelas OTHERS selama proses pengujian.

Perbedaan yang cukup besar antara nilai F1-score macro pada skenario biner dan skenario tiga kelas terlihat secara jelas pada hasil evaluasi. Pada klasifikasi biner, nilai F1-score macro berada di atas 96%, sedangkan pada klasifikasi tiga kelas nilainya berada pada kisaran 64–65%. Penurunan tersebut terjadi karena perhitungan macro average memberikan bobot yang sama pada seluruh kelas. Dengan nilai F1-score kelas OTHERS sebesar 0,00%, rata-rata F1-score

keseluruhan menjadi menurun meskipun performa pada kelas SALAH dan PENIPUAN tetap tinggi. Oleh karena itu, nilai F1-score macro pada skenario tiga kelas menjadi jauh lebih rendah dibandingkan skenario klasifikasi biner.

Jika dibandingkan secara keseluruhan, performa terbaik dalam penelitian ini diperoleh pada skenario klasifikasi biner menggunakan Focal Loss. Model tersebut menghasilkan nilai akurasi sebesar 98,63% dan F1-score macro sebesar 98,33%, yang merupakan nilai tertinggi di antara seluruh skenario pengujian. Sementara itu, pada skenario klasifikasi tiga kelas, model Cross Entropy memperoleh hasil yang sedikit lebih baik dibandingkan model Focal Loss dengan akurasi sebesar 96,64% dan F1-score macro sebesar 65,06%. Hasil tersebut menunjukkan bahwa performa model dipengaruhi oleh skenario klasifikasi yang digunakan. Perbedaan performa juga terlihat dari kemampuan model dalam mengenali seluruh kelas yang tersedia pada masing-masing skenario.

Secara umum, hasil penelitian menunjukkan bahwa model IndoBERT mampu melakukan klasifikasi berita hoaks dengan tingkat akurasi yang tinggi. Pada skenario klasifikasi biner, kedua model berhasil menghasilkan performa yang sangat baik dengan nilai F1-score di atas 95% pada seluruh kelas. Pada skenario klasifikasi tiga kelas, performa klasifikasi untuk kelas SALAH dan PENIPUAN tetap tinggi, namun nilai F1-score macro mengalami penurunan karena kelas OTHERS tidak berhasil dikenali oleh model. Berdasarkan seluruh hasil pengujian yang telah dilakukan, kombinasi IndoBERT dan Focal Loss memberikan hasil terbaik pada skenario klasifikasi biner, sedangkan kombinasi

IndoBERT dan Cross Entropy memberikan hasil terbaik pada skenario klasifikasi tiga kelas.

### 4.4 Integrasi Islam

Dalam ajaran Islam, aktivitas menerima, memverifikasi, dan menyebarkan informasi merupakan bagian dari tanggung jawab moral setiap individu. Informasi yang tidak benar atau tidak diverifikasi berpotensi menimbulkan fitnah, kesalahpahaman, perpecahan, serta kerugian bagi diri sendiri maupun masyarakat. Oleh karena itu, Islam mengajarkan prinsip tabayyun, yaitu melakukan klarifikasi atau verifikasi terhadap setiap informasi sebelum diterima dan disebarluaskan. Prinsip tersebut tidak hanya mencerminkan ketaatan seorang hamba kepada Allah SWT, tetapi juga menjadi landasan dalam menjaga hubungan yang baik antarsesama manusia. Dalam konteks penelitian ini, pengembangan sistem klasifikasi berita hoaks berbasis IndoBERT dengan pendekatan Cross Entropy Loss dan Focal Loss merupakan salah satu bentuk pemanfaatan teknologi kecerdasan buatan sebagai wasilah (sarana) untuk membantu proses identifikasi informasi yang berpotensi hoaks. Meskipun demikian, hasil klasifikasi yang diberikan oleh model tidak dapat dijadikan sebagai penentu kebenaran secara mutlak, melainkan sebagai alat bantu yang tetap memerlukan verifikasi lebih lanjut oleh manusia. Berdasarkan hal tersebut, pembahasan perspektif Islam pada penelitian ini ditinjau melalui dua aspek, yaitu Muamalah ma'a Allah (hablun minallah) sebagai hubungan manusia dengan Allah SWT dan Muamalah ma'a an-Nas (hablun minannas) sebagai hubungan manusia dengan sesama manusia.

### 4.4.1 Perspektif Muamalah ma'a Allah (Hablun Minallah

Dalam ajaran Islam, aktivitas menerima, memverifikasi, dan menyebarkan informasi merupakan bagian dari tanggung jawab moral setiap individu. Informasi yang tidak benar atau tidak diverifikasi berpotensi menimbulkan fitnah, kesalahpahaman, perpecahan, serta kerugian bagi diri sendiri maupun masyarakat. Oleh karena itu, Islam mengajarkan prinsip tabayyun, yaitu melakukan klarifikasi atau verifikasi terhadap setiap informasi sebelum diterima dan disebarluaskan. Prinsip tersebut tidak hanya mencerminkan ketaatan seorang hamba kepada Allah SWT, tetapi juga menjadi landasan dalam menjaga hubungan yang baik antarsesama manusia. Dalam perspektif Muamalah ma'a Allah (hablun minallah), penerapan prinsip tabayyun dalam menerima dan menyebarkan informasi merupakan bentuk ketaatan kepada perintah Allah SWT. Allah SWT berfirman dalam QS. Al-Hujurat ayat 6.

يَٰٓأَيُّهَا ٱلَّذِينَ ءَامَنُوٓاْ إِن جَآءَكُمْ فَاسِقٌۢ بِنَبَإٖ فَتَبَيَّنُوٓاْ أَن تُصِيبُواْ قَوْمَۢا بِجَهَٰلَةٖ فَتُصْبِحُواْ عَلَىٰ مَا فَعَلْتُمْ نَٰدِمِينَ

"*Wahai orang-orang yang beriman! Jika seseorang yang fasik datang kepadamu membawa suatu berita, maka telitilah kebenarannya agar kamu tidak mencelakakan suatu kaum karena kebodohan (kecerobohan), yang akhirnya kamu menyesali perbuatanmu itu*." (QS. Al-Hujurat [49]: 6).

Ayat tersebut memerintahkan umat Islam untuk memverifikasi setiap informasi sebelum mempercayai maupun menyebarkannya agar tidak menimbulkan penyesalan akibat kesalahan dalam mengambil keputusan. Selain itu, Allah SWT juga berfirman dalam QS. An-Nur ayat 11:

إِنَّ ٱلَّذِينَ جَآءُو بِٱلْإِفْكِ عُصْبَةٌ مِّنكُمْ ۚ لَا تَحْسَبُوهُ شَرًّا لَّكُم ۖ بَلْ هُوَ خَيْرٌ لَّكُمْ ۚ لِكُلِّ ٱمْرِئٍ مِّنْهُم مَّا ٱكْتَسَبَ مِنَ ٱلْإِثْمِ ۚ وَٱلَّذِى تَوَلَّىٰ كِبْرَهُۥ مِنْهُمْ لَهُۥ عَذَابٌ عَظِيمٌ

"*Sesungguhnya orang-orang yang membawa berita bohong itu adalah dari golongan kamu juga. Janganlah kamu mengira bahwa berita bohong itu buruk bagi kamu, bahkan itu baik bagi kamu. Setiap orang dari mereka mendapat balasan dari dosa yang diperbuatnya. Dan barang siapa di antara mereka yang mengambil bagian terbesar dalam penyiaran berita bohong itu, baginya azab yang besar*." (QS. An-Nur [24]: 11).

Menurut Tafsir Ibnu Katsir (Ibnu Katsir, 2000), ayat tersebut turun berkaitan dengan peristiwa fitnah terhadap Sayyidah Aisyah RA yang mengajarkan pentingnya kehati-hatian dalam menerima dan menyampaikan informasi. Senada dengan itu, Al-Qurthubi (2003) menjelaskan bahwa istilah al-ifki bermakna kebohongan besar yang memutarbalikkan fakta.

Prinsip tersebut juga diperkuat oleh sabda Rasulullah SAW:

كَفَىٰ بِالْمَرْءِ كَذِبًا أَنْ يُحَدِّثَ بِكُلِّ مَا سَمِعَ

"*Cukuplah seseorang dikatakan berdusta apabila ia menceritakan setiap apa yang didengarnya*." (HR. Muslim No. 5).

Hadis tersebut menunjukkan bahwa setiap muslim memiliki tanggung jawab moral untuk memverifikasi informasi sebelum menyampaikannya kepada orang lain. Dalam konteks penelitian ini, sistem klasifikasi berita hoaks berbasis IndoBERT merupakan salah satu bentuk ikhtiar yang dapat membantu proses tabayyun dengan mengidentifikasi potensi berita hoaks secara otomatis. Meskipun demikian, hasil klasifikasi tidak dapat dijadikan penentu kebenaran secara mutlak karena model hanya mempelajari pola bahasa dari data dan tetap memerlukan verifikasi oleh manusia. Dengan demikian, teknologi kecerdasan buatan berfungsi

sebagai wasilah (sarana) yang mendukung pelaksanaan perintah Allah SWT dalam menjaga kejujuran informasi sekaligus mendukung tujuan Maqāṣid al-Syarī'ah, khususnya menjaga akal (ḥifẓ al-'aql) dari pengaruh informasi yang menyesatkan.

### 4.4.2 Perspektif Muamalah ma'a an-Nas (Hablun Minannas)

Dalam ajaran Islam, aktivitas menerima, memverifikasi, dan menyebarkan informasi merupakan bagian dari tanggung jawab moral setiap individu. Informasi yang tidak benar atau tidak diverifikasi berpotensi menimbulkan fitnah, kesalahpahaman, perpecahan, serta kerugian bagi diri sendiri maupun masyarakat. Oleh karena itu, Islam mengajarkan prinsip tabayyun, yaitu melakukan klarifikasi atau verifikasi terhadap setiap informasi sebelum diterima dan disebarluaskan. Prinsip tersebut tidak hanya mencerminkan ketaatan seorang hamba kepada Allah SWT, tetapi juga menjadi landasan dalam menjaga hubungan yang baik antarsesama manusia. Dalam perspektif Muamalah ma'a Allah (hablun minallah), penerapan prinsip tabayyun dalam menerima dan menyebarkan informasi Dalam perspektif Muamalah ma'a an-Nas (hablun minannas), penyebaran informasi yang tidak benar dapat menimbulkan fitnah, kesalahpahaman, perselisihan, hingga konflik sosial yang merusak hubungan antarsesama manusia. Islam mengajarkan agar setiap muslim menjaga kejujuran dalam komunikasi demi terciptanya kehidupan sosial yang harmonis. Rasulullah SAW bersabda:

إِنَّ كَذِبًا عَلَيَّ لَيْسَ كَكَذِبٍ عَلَى أَحَدٍ، مَنْ كَذَبَ عَلَيَّ مُتَعَمِّدًا، فَلْيَتَبَوَّأْ مَقْعَدَهُ مِنَ النَّارِ

"*Sesungguhnya berdusta atas namaku tidak sama dengan berdusta atas nama orang lain. Barang siapa yang sengaja berdusta atas namaku, maka hendaklah ia menyiapkan tempat duduknya di neraka*." (HR. Bukhari No. 1291 dan HR. Muslim No. 3).

Hadis tersebut menegaskan bahwa penyampaian informasi yang tidak benar merupakan perbuatan yang sangat dilarang dalam Islam karena dapat menimbulkan kerugian bagi orang lain. Sejalan dengan ajaran tersebut, penelitian ini mengembangkan sistem klasifikasi berita hoaks berbasis IndoBERT menggunakan pendekatan Cross Entropy Loss dan Focal Loss untuk membantu mengidentifikasi berita hoaks secara lebih akurat, khususnya pada kelas minoritas. Sistem ini diharapkan dapat meminimalkan penyebaran informasi palsu sehingga mampu mengurangi potensi fitnah, keresahan, dan konflik di masyarakat. Dengan demikian, teknologi kecerdasan buatan dapat dimanfaatkan sebagai wasilah dalam menjaga hubungan antarsesama manusia (hablun minannas), mendukung terciptanya ketertiban sosial, serta mewujudkan kemaslahatan bersama sesuai dengan nilai-nilai Islam.

# BAB V

# KESIMPULAN DAN SARAN

## 5.1 Kesimpulan

Berdasarkan hasil analisis dan eksperimen yang telah dilakukan pada penelitian ini, dapat ditarik kesimpulan sebagai berikut:

a. Model IndoBERT berhasil diimplementasikan untuk melakukan klasifikasi biner konten hoaks pada kategori SALAH dan PENIPUAN dengan performa yang sangat baik. Pada skenario baseline menggunakan Cross Entropy Loss, model memperoleh akurasi sebesar 97,26% dengan F1-score macro sebesar 96,56%, di mana F1-score untuk kelas SALAH mencapai 98,11% dan kelas PENIPUAN mencapai 95,00%. Hasil tersebut menunjukkan bahwa IndoBERT mampu memahami konteks bahasa Indonesia secara efektif sehingga dapat membedakan karakteristik berita hoaks kategori SALAH dan PENIPUAN dengan tingkat akurasi yang tinggi. Selain itu, penggunaan preprocessing minimal juga terbukti sesuai dengan karakteristik model berbasis Transformer karena mampu mempertahankan informasi linguistik yang penting dalam proses klasifikasi.
b. Penerapan teknik penanganan class imbalance menggunakan Focal Loss dengan parameter $\alpha = 0,75$ dan $\gamma = 2,0$ terbukti mampu meningkatkan performa model IndoBERT, terutama dalam mendeteksi kelas minoritas PENIPUAN. Model yang menggunakan Focal Loss memperoleh akurasi sebesar 98,63% dan F1-score macro sebesar 98,33%, lebih tinggi

dibandingkan pendekatan Cross Entropy Loss. Peningkatan paling signifikan terjadi pada kelas PENIPUAN dengan F1-score yang meningkat dari 95,00% menjadi 97,62%, sedangkan nilai loss menurun dari 0,0923 menjadi 0,0059 atau turun sebesar 93,61%. Hasil confusion matrix juga menunjukkan bahwa penggunaan Focal Loss berhasil menghilangkan seluruh kesalahan false negative pada kelas SALAH, sehingga model menjadi lebih optimal dalam mengenali kedua kelas. Dengan demikian, penelitian ini membuktikan bahwa penerapan Focal Loss efektif dalam mengatasi ketidakseimbangan data dan meningkatkan kemampuan model IndoBERT dalam mendeteksi kelas minoritas tanpa menurunkan performa klasifikasi secara keseluruhan. Dalam perspektif Islam, hasil penelitian ini sejalan dengan prinsip tabayyun sebagaimana diperintahkan dalam QS. Al-Hujurat ayat 6, yaitu pentingnya memverifikasi kebenaran informasi sebelum menerima maupun menyebarluaskannya. Penggunaan teknologi kecerdasan buatan untuk mendeteksi berita hoaks juga mendukung tujuan Maqāṣid al-Syarī‘ah, khususnya dalam menjaga akal (ḥifẓ al-'aql) dan memelihara keharmonisan sosial dari dampak negatif penyebaran informasi palsu.

## 5.2 Saran

Peneliti menyadari bahwa hasil dari penelitian ini masih memiliki keterbatasan dan belum sepenuhnya sempurna. Agar sistem yang dikembangkan dapat berfungsi secara lebih optimal dan memberikan hasil yang lebih komprehensif, diperlukan beberapa langkah pengembangan lanjutan yang dapat

meningkatkan kualitas dan performa analisis. Adapun saran untuk penelitian selanjutnya adalah sebagai berikut:

1. Mengambil data dari berbagai platform berita atau media sosial untuk memperkaya konteks dan keragaman pola hoaks, sehingga model dapat belajar dari lebih banyak variasi teks.
2. Eksplorasi hyperparameter model dapat diperdalam, tidak terbatas pada parameter alpha dan gamma pada Focal Loss, tetapi juga mencakup learning rate, batch size, jumlah epoch, arsitektur model, dan strategi regularization tambahan untuk mengidentifikasi konfigurasi yang paling optimal.
3. Membandingkan performa IndoBERT dengan model pre-trained bahasa Indonesia lainnya seperti IndoBERT-large, RoBERTa-base-Indo, atau DistilBERT untuk mengetahui model mana yang paling efisien dan akurat dalam tugas klasifikasi berita hoaks.
4. Mengintegrasikan anotasi tambahan seperti tingkat hoaks (misalnya hoaks ringan, sedang, berat) atau kategori topik hoaks (politik, kesehatan, ekonomi) untuk memperluas dimensi analisis dan menghasilkan klasifikasi yang lebih granular.

# DAFTAR PUSTAKA


Abidin, D. Z., Siswanto, A., Saputra, C., Betantiyo, B., & Nehemia Toscany, A. (2025). Enhancing Fake News Detection on Imbalanced Data Using Resampling Techniques and Classical Machine Learning Models. *Jurnal Teknik Informatika (Jutif)*, *6*(5), 3769–3786. https://doi.org/10.52436/1.jutif.2025.6.5.5177

Aslan, L., Ptaszynski, M., & Jauhiainen, J. (2024). Are Strong Baselines Enough? False News Detection with Machine Learning. *Future Internet*, *16*(9), 322. https://doi.org/10.3390/fi16090322

Asshiddiq, Muh. H., & Witanti, A. (2025). Analisis Sentimen tentang Penundaan Pengangkatan CPNS 2025 pada Platform X Menggunakan Metode IndoBERT. *Jurnal teknika*, *17*(2), 97–108. https://doi.org/10.30736/jt.v17i2.1448

Beseiso, M., & Al-Zahrani, S. (2025). A Context-Enhanced Model for Fake News Detection. *Engineering, Technology & Applied Science Research*, *15*(1), 19128–19135. https://doi.org/10.48084/etasr.9192

Dai, L., Zhou, M., & Liu, H. (2023). Recent Applications of Convolutional Neural Networks in Medical Data Analysis: Dalam A. Hassan, V. K. Prasad, P. Bhattacharya, P. Dutta, & R. Damaševičius (Ed.), *Advances in Healthcare Information Systems and Administration* (hlm. 119–131). IGI Global. https://doi.org/10.4018/979-8-3693-1082-3.ch007

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (t.t.). *BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding*.

Dhendra, & Gayuh Utomo, V. (2025). Benchmarking IndoBERT and Transformer Models for Sentiment Classification on Indonesian E-Government Service Reviews. *Jurnal Transformatika*, *23*(1), 86–95. https://doi.org/10.26623/transformatika.v23i1.12095

Faisal, D. R., & Mahendra, R. (2022). *Two-Stage Classifier for COVID-19 Misinformation Detection Using BERT: A Study on Indonesian Tweets* (arXiv:2206.15359). arXiv. https://doi.org/10.48550/arXiv.2206.15359

Gao, T., Yao, X., & Chen, D. (2022). *SimCSE: Simple Contrastive Learning of Sentence Embeddings* (arXiv:2104.08821). arXiv. https://doi.org/10.48550/arXiv.2104.08821

Hu, B., Mao, Z., & Zhang, Y. (2025). An overview of fake news detection: From a new perspective. *Fundamental Research*, *5*(1), 332–346. https://doi.org/10.1016/j.fmre.2024.01.017

Koto, F., Lau, J. H., & Baldwin, T. (2021). IndoBERTweet: A Pretrained Language Model for Indonesian Twitter with Effective Domain-Specific Vocabulary Initialization. *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, 10660–10668. https://doi.org/10.18653/v1/2021.emnlp-main.833

Lin, T.-Y., Goyal, P., Girshick, R., He, K., & Dollar, P. (t.t.). *Focal Loss for Dense Object Detection*.

Loshchilov, I., & Hutter, F. (2019). *Decoupled Weight Decay Regularization* (arXiv:1711.05101). arXiv. https://doi.org/10.48550/arXiv.1711.05101

Munir, A. (2020). *GERAKAN PEMUDA CERDAS LAWAN BERITA HOAX*. *01*(01).

Prisscilya, V., & Girsang, A. S. (2024). Classification of Indonesia False News Detection Using Bertopic and Indobert. *Jurnal Indonesia Sosial Teknologi*, *5*(8), 3061–3079. https://doi.org/10.59141/jist.v5i8.1310

Putra, H. K., Bijaksana, M. A., & Romadhony, A. (t.t.). *Deteksi Penggunaan Kalimat Abusive Pada Teks Bahasa Indonesia Menggunakan Metode IndoBERT*.

Rahadi, D. R. (2017). PERILAKU PENGGUNA DAN INFORMASI HOAX DI MEDIA SOSIAL. *JURNAL MANAJEMEN DAN KEWIRAUSAHAAN*, *5*(1). https://doi.org/10.26905/jmdk.v5i1.1342

Ridho, M. Y., & Yulianti, E. (2024). From Text to Truth: Leveraging IndoBERT and Machine Learning Models for Hoax Detection in Indonesian News. *Jurnal Ilmiah Teknik Elektro Komputer Dan Informatika*, *10*(3), 544–555. https://doi.org/10.26555/jiteki.v10i3.29450

Rona, A. R., Syarifudin, A., & Hamandia, M. R. (2024). Komunikasi Islam dalam Pembinaan Praktek Ibadah Kemasyarakatan pada Jama'ah Masjid Muqoddimatul Hidayah Talang Kelapa Palembang. *Jurnal Pendidikan Islam*, *1*(4), 8. https://doi.org/10.47134/pjpi.v1i4.853

Roumeliotis, K. I., Tselikas, N. D., & Nasiopoulos, D. K. (2025). Fake News Detection and Classification: A Comparative Study of Convolutional Neural Networks, Large Language Models, and Natural Language Processing Models. *Future Internet*, *17*(1), 28. https://doi.org/10.3390/fi17010028

Sanh, V., Debut, L., Chaumond, J., & Wolf, T. (2020). *DistilBERT, a distilled version of BERT: Smaller, faster, cheaper and lighter* (arXiv:1910.01108). arXiv. https://doi.org/10.48550/arXiv.1910.01108

Tala, F. Z. (t.t.). *A Study of Stemming Effects on Information Retrieval in Bahasa Indonesia*.

Yefferson, D. Y., Lawijaya, V., & Girsang, A. S. (2024). Hybrid model: IndoBERT and long short-term memory for detecting Indonesian hoax news. *IAES International Journal of Artificial Intelligence (IJ-AI)*, *13*(2), 1913. https://doi.org/10.11591/ijai.v13.i2.pp1913-1924